Software Engineering Institute

Approaches to Process Performance Modeling: A Summary from the SEI Series of Workshops on CMMI High Maturity Measurement and Analysis

Robert W. Stoddard II
Dennis R. Goldenson

January 2010

http://www.sei.cmu.edu

Carnegie Mellon

Table of Contents

List of Figures

List of Tables

Acknowledgments

Abstract

More and more organizations are striving for and achieving high maturity status, yet there is still an insufficient shared understanding of how best to implement measurement and analysis practices appropriate for high maturity organizations. A series of twice-yearly workshops organized by the Software Engineering Institute (SEI) allows organizations to share lessons learned to accelerate the adoption of best measurement and analysis practices in high maturity organizations.

This report summarizes the results from the second and third high maturity measurement and analysis workshops. The participants' presentations described their experiences with process performance models; the goals and outcomes of the modeling; the x factors used; the data collection methods; and the statistical, simulation, or probabilistic modeling techniques used. Overall summaries of the experience and future plans for modeling also were provided by participants.

This report also includes a summary of the "healthy ingredients" that are needed for process performance models and a table showing which healthy ingredients were visible in the models described in the presentations. By making the models that were shared in these workshops more widely available in this report, the community as a whole can benefit from the exciting and innovative ideas for process performance models implemented by leading organizations in the field.

1 Introduction

More and more organizations are striving for and achieving high maturity status, yet there is still an insufficient shared understanding of how best to implement measurement and analysis practices that are appropriate for high maturity organizations. A series of twice-yearly workshops organized by the SEI allows organizations to share lessons learned to accelerate the adoption of best measurement and analysis practices in high maturity organizations.

The workshops provide a unique opportunity for candid discussions among participants from leading organizations in the field. Participants from different units of the same large organizations who had not known each other before have begun to work together collaboratively. People from different organizations have also begun to work together to their mutual benefit. The broader goal of the workshop series is to promote the establishment of a viable community of interest around high maturity measurement and analysis that extends well beyond the relatively small number of workshop participants.

Three workshops were held prior to the writing of this report, focusing on the deployment, adoption, and institutionalization of process performance baselines and models. As more workshops are held, the focus will broaden to include other analytic methods, how they have been used, the performance and quality outcomes that have been achieved, and the organizational circumstances under which such models are likely to be used effectively.

The workshops in this series are meant to guide ongoing research and transition activities related to process performance modeling and analytical practices. The participants discussed what they found useful for quantitative management, process performance, and product quality in their organizations, regardless of whether those practices fully satisfied CMMI high maturity appraisal criteria. Discussions of high maturity appraisal criteria and updates to the CMMI models and the SCAMPI A Method Definition Document (MDD) remained for the most part outside the scope of the workshops and this document.

1.1 Workshops Two and Three: Goals and Participants

Participation in the first workshop was limited to a small group of organizations who were early adopters of process performance models and baselines and was by invitation only. A summary of the first workshop can be found in *CMMI High Maturity Measurement and Analysis Workshop Report: March 2008* [1]. Wider solicitations for proposals were made for the second and third workshops. Participants were required to submit case descriptions with empirical evidence for review before being accepted to the workshop. The presentations needed to include thorough case studies, with results presented in quantitative terms. The practice descriptions needed to be discussed in sufficient detail to be meaningful to the other workshop participants.

Participants were asked to include the following topics in their presentations:

- descriptions of the process performance models, including who developed them, what initiated the development, who used them, and how often they were used
- business and project goals and how the goal alignment led to the identity and use of the model

- outcomes predicted by the process performance models (PPMs), the type of data output, and the audience for each outcome and why it was important to them
- descriptions of the x factors used in the model to predict the outcomes
- data collection practices
- a description of all statistical, simulation, or probabilistic modeling techniques used and the rationale for their selection
- results and benefits of using the process performance models
- an overall summary of the experience and future plans for modeling in the organization

This can be and was done productively without disclosing proprietary information. Quantitative results occasionally were normalized for the same reason.

The second workshop was held in Denver, Colorado, November 21-22, 2008. Presentations were prepared by the following:

- Stephen Austin and Bob Beckley, Lockheed Martin
- Dan Bennett, Rushby Craig, and Kevin Tjoland; Ogden Air Logistics Center
- Pedro E. Colla, Instituto Universitario Aeronáutico and Universidad Tecnológica Nacional-Facultad Regional Santa Fé; and Jorge Marcelo Montagna, INGAR-Instituto de Desarrollo y Diseño, Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional-Facultad Regional Santa Fé
- Rick Hefner, Northrop Grumman Mission Systems
- Mark Kelley, Esterline AVISTA
- Shankar Mallapur, Johnson Varghese, and Gayathri Pallail; Accenture Services Pvt Ltd
- Neal Mackertich, Michael Campo, and Rachel Beitz; Raytheon Integrated Defense Systems
- Lynn Penn and Pete McLoone, Lockheed Martin Information Systems and Global Services
- Jim Perry, BAE Armament Systems
- David M. Raffo, Portland State University
- Kobi Vider-Picker, K.V.P. Consulting
- David Webb, Hill Air Force Base; David Tuma, Tuma Solutions; Jim Van Buren, Draper Laboratory; and Robert Stoddard, Software Engineering Institute

The third workshop was held in San Jose, California, March 27-28, 2009. Presentations were prepared by the following:

- Yoshihiro Akiyama, Kyushu Institute, Technology and Next Process Institute
- Pedro E. Colla, Instituto Universitario Aeronáutico and Universidad Tecnológica Nacional-Facultad Regional Santa Fé; and Jorge Marcelo Montagna, INGAR-Instituto de Desarrollo y Diseño, Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional-Facultad Regional Santa Fé
- Brooke Eiche, Lockheed Martin Systems Integration
- Mike Konrad, Software Engineering Institute
- Angel Liu, Motorola China

- Diane Mizukami Williams, Northrop Grumman Corporation

- David M. Raffo, Portland State University

- Kathy Smith, Hewlett Packard

- Lynn Penn, Lockheed Martin Information Systems and Global Services

- Neal Mackertich and Michael Campo, Raytheon Integrated Defense Systems

- Raymond Kile, John Gaffney, and Joan Weszka; Lockheed Martin Corporate Engineering and Technology Systems and Software Resource Center

- Kobi Vider-Picker, K.V.P. Consulting

1.2 Structure of Workshops

Workshops two and three began with presentations from the participants, with time allowed after each for questions and answers.

The presentations were followed by breakout working sessions with teams meeting in parallel. The goal of the sessions was to produce recommendations related to reducing barriers to effective training and staffing, management support, alignment of modeling to business goals, and using different analytical forms of modeling. Workshop three also included a panel session on issues and obstacles to effective process performance modeling.

The presentations covered many different analytical approaches, including statistical or mathematical, descriptive, probabilistic, and simulation. Some used large-scale baselines while others used small datasets. Still others addressed issues of coping with missing and imperfect data, as well as the use of expert judgment to calibrate the models. Most of the presentations described the use of the models in large organizations consisting of multiple and sometimes disparate stakeholders, but smaller organizations also were included as were comparisons across organizations. The interim and final performance outcomes predicted by the models also differed considerably (e.g., defect prevention, customer satisfaction, other quality attributes, aspects of requirements management, return on investment, cost, schedule, efficiency of resource usage, and staff skills as a function of training practices). Of course all of the various predictive factors also differed as a function of the outcomes predicted by the models that were presented.

Thumbnail descriptions of the presentations can be seen in Table 1 beginning on page 8, along with a discussion of the ways in which the healthy ingredients of a CMMI process performance model were covered in the presentations. Fuller synopses of the presentations can be found in Section 2. They are categorized by the different analytical approaches used in the presentations.[1] The panel and break-out sessions are summarized in Section 3.

1.3 Healthy Ingredients of Process Performance Models

Discussion of the "healthy ingredients" of CMMI process performance models began in 2007 with SEI presentations at SEPG conferences and was amplified in the SEI course Understanding CMMI High Maturity Practices (UCHMP). The healthy ingredients were first elaborated dynami-

---

[1] An additional synopsis of a presentation that briefly discussed analytical methods aimed at "Improving Estimates of Prediction Error" also appears in Section 3.

cally during the conduct of SEI measurement courses in 2006 and 2007 as a means of communicating what process performance models were in concrete, practical terms. The ingredients are derived from a holistic understanding of the intent of the CMMI models. The precise nature of several of the ingredients also comes from training, experience, and practice within the Six Sigma arena. The healthy ingredients of a process performance model are summarized below.

- The model is statistical, probabilistic, or simulation based. This particular ingredient emphasizes the logical consistency of two CMMI process areas: Quantitative Project Management (QPM) and Organizational Process Performance (OPP). QPM stresses the need for understanding statistical variation of process performance factors. Additionally, QPM reinforces the need to separate assignable, special cause variation from inherent common cause variation to help understand what actions to take with respect to each type of variation. This healthy ingredient emphasizes the need for process performance models to model the uncertainty of the predictive factors and their resulting impact on the uncertainty of the behavior of the outcome factor. For this reason, deterministic models that merely perform mathematical calculations on point estimates fall short of the superior information achievable from models that are statistical, probabilistic, or simulation in nature.

- The model predicts interim and/or final project outcomes. This ingredient derives more from practical experience and management's need for real-time cycles of learning within a given project or program. To maximize real-time cycles of learning within a given project or program, managers need to predict interim performance outcomes in addition to the traditional end-of-project performance outcomes.

- The model uses controllable predictive factors that are directly tied to subprocesses or work activities. This healthy ingredient focuses on the need for process performance models to be actionable. From that standpoint, if a model does not have at least one controllable predictive factor, it does not directly promote insight for action to influence the undesirable predicted outcome. For example, traditional project forecasting models that model only uncontrollable factors make predictions that offer little help or insight into the actions to be taken to drive a more desirable predicted outcome. Additionally, this ingredient highlights the need for the controllable factors to be detailed enough to show a clear link to a specific subprocess or work activity. This clear link enables proactive management responses.

- The model quantitatively characterizes and models the variation of the predictive factors and describes the predicted range, uncertainty, or variation of the outcome performance measures. This ingredient is a chief overlap of CMMI high maturity and Six Sigma concepts. Recognizing that variation (i.e., risk) may very well be unbalanced and significant in the real world, the models account for this by modeling the uncertainty of the predictive factors. Numerous examples exist in industry in which analysis using only the mean or average estimate rather than the distributional information caused serious problems in predictions of schedule, performance, and other modeled factors.

- The model enables "what-if" analysis for project planning, dynamic re-planning, and problem resolution during project execution. This ingredient builds on language in the CMMI Organizational Process Performance (OPP), Quantitative Project Management (QPM), Organizational Innovation and Deployment (OID), and Causal Analysis and Resolution (CAR) process areas related to the use of process performance models to support "what-if" and sensitivity analysis. The idea is that decision makers will be able to use process performance models to

analyze alternative courses of action and alternative improvement ideas. Again, this high-lights a capability intended to be exercised within a given project or program execution.

- The model connects upstream activity with downstream activity. This particular ingredient emphasizes the intent of process performance models to enable decision makers to observe a prediction of the consequences of decisions made earlier in the life cycle or process. Indeed, this ingredient highlights the practical use of process performance models for transitions from phase to phase, hand-offs from one group to another, and so on. This particular ingredient enables the establishment and enforcement of interface agreements between internal groups and/or external groups by providing models that predict the readiness and maturity of an artifact or work product to proceed to the next step. For example, many organizations employ such models to predict defects entering system test while the code is still with the development team. Others use models to predict readiness of design or code to enter an inspection. Still other organizations use models in this fashion to determine if product and software requirements are sufficiently mature and stable to begin intense development.

- The model enables projects to achieve mid-course corrections to ensure project success. This ingredient highlights a very significant aspect that may be read into the usage of process performance models in CMMI. Specifically, within the QPM process area, process performance models may be used to anticipate undesirable performance with enough lead time to proactively influence the situation toward a successful outcome. Industry experience with this aspect is quite strong, especially in the use of critical parameter management in the Design for Six Sigma (DFSS) community. The notion is that models of critical parameters of the product design foster early insight into issues in products and processes enabling management to take corrective and preventive action. For this reason, organizations employ a collection of process performance models to cover their needs throughout the project life cycle.

Figure 1 expresses the healthy ingredients visually. The outermost box represents all models, be they qualitative or quantitative in nature. The innermost box represents the set of models that faithfully meet the entire list of healthy ingredients. While they may still offer substantial business value, the outer boxes moving inward describe other types of modeling that may include only one or more of the healthy ingredients. To read the figure, focus first on the outermost box labeled "All Models (Qualitative and Quantitative)". The "Anecdotal Biased Samples" label on the right side of the box emphasizes that the set of all models that reside in the outermost layer may contain only anecdotal biased samples. Similarly, models represented by the first embedded box, "Quantitative Models (Deterministic, Statistical, Probabilistic)," go beyond anecdotal biased samples and are based on quantitative data; however, they may not model uncertainty or variation. Models represented by each subsequent box embedded inward add a new healthy ingredient as they continue to incorporate the sets of healthy ingredients that were missing previously.
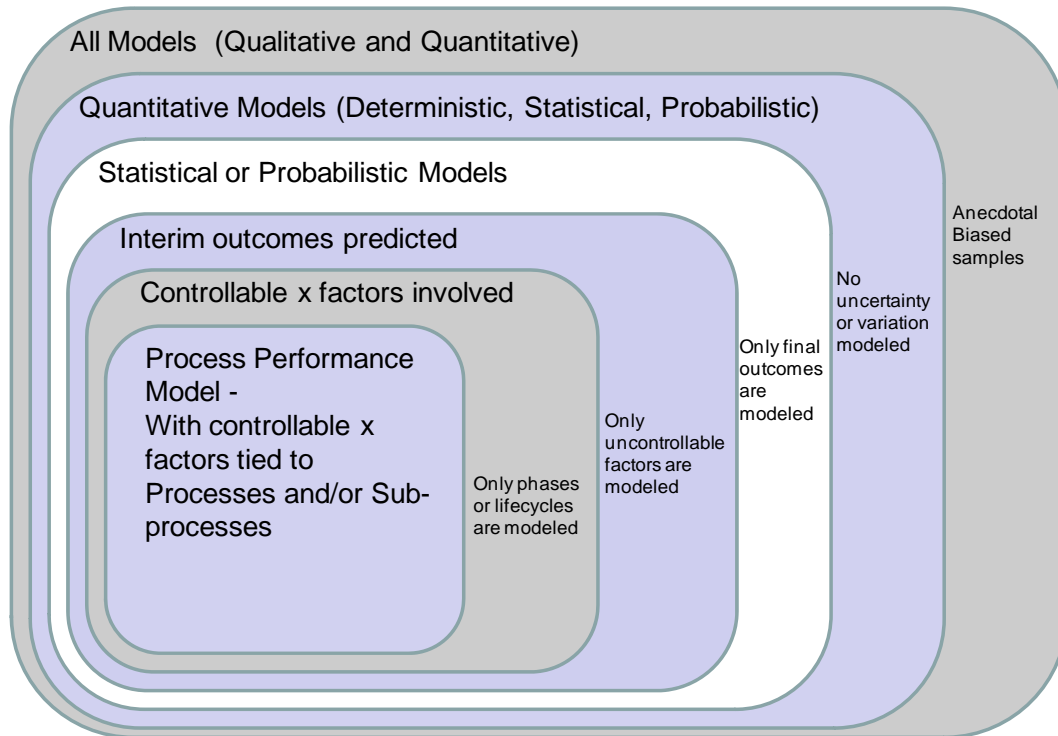
Figure 1:   Healthy Ingredients of Process Performance Models

The discussion of the healthy ingredients in this technical report comes from a perspective of both the intent of the CMMI model and the practical use and benefits now apparent in industry. The discussion purposely excludes commentary related to the SCAMPI appraisal perspective and does not purport to address the criteria used by the SCAMPI method and lead appraiser community for acceptable process performance model implementation.

With that said, the workshop series on CMMI high maturity measurement and analysis was designed to communicate the healthy ingredients included and the business benefits gained by the workshop participants in using models that strive to meet the definition and purpose of process performance models. Not all models in practice embody the entire set of healthy ingredients. However, organizations have used collections of process performance models that together achieve the set of healthy ingredients. This positive discussion is also meant to promote additional innovative thought as to how the existing models might be further enhanced with additional treatment of other healthy ingredients.

By making the models that were shared in these two workshops more widely available in this report, we hope that the community as a whole will benefit from the exciting and innovative ideas for process performance models implemented by leading organizations in the field.

Table 1 identifies the healthy ingredients that were clearly evident to the SEI workshop hosts during the individual workshop presentations. Each presentation discussed some or all of the healthy ingredients. In some cases, the presentation authors acknowledged that future work would be pursued to incorporate missing healthy ingredients. In other cases, the presentation authors did not discuss all of the healthy ingredients or chose, for business purposes, not to include all of them in their models. As such, this table is not intended to grade the models, but instead to aid the reader in quickly searching for models that emphasize particular healthy ingredients in more detail.

The healthy ingredients of the models are summarized in the table on the next page. The numbers indicate which healthy ingredients were visible in each model or models described in each presentation. The table is categorized by the analytical approaches used in the presentations, which are listed in the same order as the fuller descriptions in Section 2.

Table 1:   Healthy Ingredients in Presentations

1. was statistical, probabilistic, or simulation in nature
2. predicted interim and/or final project outcomes
3. used controllable predictive factors directly tied to subprocesses or work activities
4. quantitatively characterized and modeled the variation of the predictive factors and understood the predicted range, uncertainty, or variation of the outcome performance measures
5. enabled what-if analysis for project planning, dynamic re-planning, and problem resolution during project execution
6. connected upstream activity with downstream activity
7. enabled projects to achieve mid-course corrections to ensure project success

Presentation	Presenter	Model Purpose	1	2	3	4	5	6	7
Discrete Event Simulation									
Putting Process Performance Models to Work in the Real World: An Example and Recommendations *(p 11)*	Raffo	Used discrete event simulation to predict schedule variance and quality based on a variety of life-cycle controllable factors.	X	X	X	X	X	X	X
Putting Process Performance Models to Work: A NASA Example *(p 13)*	Raffo	Used discrete event simulation to predict the cost, schedule, and quality impacts of adopting the QuARs requirements analysis tool.	X	X	X	X	X	X	X
A Process Performance Model Used Within Lockheed Martin MS2 *(p 16)*	Austin	Used discrete event simulation to model the drawing control and release process so that cycle time, bottlenecks, and resource contentions could be identified and resolved as early as possible.	X	X	X	X	X	X	X
Monte Carlo									
Compose Process Model *(p 19)*	Mallapur	Used Monte Carlo simulation to predict defects and cycle time based on controllable factors of development and peer review.	X	X	X	X	X	X	X
Scheduling Analysis of Variability Engine (SAVE) *(p 23)*	Mackertich	Used Monte Carlo simulation to predict various schedule success outcomes based on uncertainty of inputs to individual work tasks in the program.	X	X		X		X	X
Process Performance Models: Process, Results, and Lessons Learned with SLAM *(p 26)*	Mackertich	Used multivariable linear regression and Monte Carlo simulation to predict cost and schedule performance based on requirements volatility and the degree of overlap of the requirements and design phases (e.g., surrogate for risk of proceeding with development prematurely)	X		X	X	X	X	X
Evaluation of SPI Efforts in Small and Medium Organizations *(p 28)*	Colla	Used both a deterministic model based on mathematical transfer functions and stochastic (Monte Carlo) scenario modeling to predict the likelihood of return on investment from the adoption of CMMI-based process improvement initiatives in small and medium software organizations in Argentina.	X	X		X	X		
Evaluation of SPI Efforts in Small and Medium Organizations: An Update with New Results *(p 31)*	Colla	Models additional Monte Carlo scenarios based on further model calibration to predict the likelihood of return on investment from the adoption of CMMI-based process improvement initiatives.	X	X		X	X		

1. was statistical, probabilistic, or simulation in nature
2. predicted interim and/or final project outcomes
3. used controllable predictive factors directly tied to subprocesses or work activities
4. quantitatively characterized and modeled the variation of the predictive factors and understood the predicted range, uncertainty, or variation of the outcome performance measures
5. enabled what-if analysis for project planning, dynamic re-planning, and problem resolution during project execution
6. connected upstream activity with downstream activity
7. enabled projects to achieve mid-course corrections to ensure project success

Presentation	Presenter	Model Purpose	1	2	3	4	5	6	7
Other Simulation Approaches									
Conceptual Planning, Execution, and Operation of Combat Fire Support Effectiveness *(p 37)*	Vider-Picker	Demonstrated the use of Bayesian belief networks and process flow simulation models for the definition of end-to-end life-cycle processes requiring coordination among disparate stakeholder groups to meet product quality objectives and efficiency of resource usage.	X	X	X	X	X	X	X
Game Theory and Bayesian Belief Network to Support QFD for Requirements Development *(p 42)*	Vider-Picker	Used Game Theory and Bayesian methods to predict program success based primarily on stakeholder and engineering factors.	X	X		X	X		
Product Modeling to Ensure Reliability of High Maturity Indicators *(p 45)*	Perry	Used algebraic, probabilistic, and Eigenvalue structure matrices to enable the prediction of impacts on software quality attributes as a result of changes to detailed subprocess tightly coupled to software and system components.	X	X	X		X	X	X
Software Maturity Modeling *(p 48)*	Penn	Used software reliability growth models to predict ship readiness using attributes and results of testing.	X	X		X			X
Defect Discovery *(p 50)*	Penn	Used Rayleigh curve fitting to predict defect discovery (e.g., depicted as defect densities by phase) across the life cycle and to also predict latent or escaping defects.	X	X		X	X	X	
Process Performance Modeling with Parametric Cost Models *(p 53)*	Kile	Used traditional cost models with modifications based on several of the healthy ingredients to predict cost, schedule, and quality.	X	X	X	X	X		X
Process Performance Baselines for Northrop Grumman Mission Systems *(p 59)*	Hefner	Used historical baselines to predict productivity and defects (injection, detection, removal, and escape).		X	X	X	X	X	X
Statistical Models									
CAPRE Effort Estimation and Tracking Model Using Dummy Variable Linear Regression *(p 60)*	Webb	Used ANOVA and dummy variable regression to predict effort and schedule by phase using a variety of controllable process factors.	X	X	X	X	X		X

1. was statistical, probabilistic, or simulation in nature
2. predicted interim and/or final project outcomes
3. used controllable predictive factors directly tied to subprocesses or work activities
4. quantitatively characterized and modeled the variation of the predictive factors and understood the predicted range, uncertainty, or variation of the outcome performance measures
5. enabled what-if analysis for project planning, dynamic re-planning, and problem resolution during project execution
6. connected upstream activity with downstream activity
7. enabled projects to achieve mid-course corrections to ensure project success

Presentation	Presenter	Model Purpose	1	2	3	4	5	6	7
Statistical Models, continued									
Customer Satisfaction (p 63)	Penn	Used multiple linear regression to predict award fees (surrogate for customer satisfaction) as a function of a number of controllable program factors.	X	X		X	X	X	
Tailoring Baselines and Models (p 65)	Mizukami Williams	Used multivariable regression analysis to predict the tailoring hours needed for specific programs based on factors including the type of tailoring method.	X	X	X	X	X		
Continuous Improvement for Cost of Quality (p 72)	Kelley	Used simple linear regression to predict the differences in Cost of Quality based on process inputs from peer reviews and product attributes.	X	X	X	X	X	X	X
Training Improvement through Competence Measurement and Analysis (p 74)	Liu	Used simple linear regression and ordinal logistic regression to predict individual manager competency and performance based on deployment of improved processes for training, mentoring, and other professional development activities.	X		X	X	X		
Program Staff Effectiveness Model (p 76)	Eiche	Used multiple variable regression to predict program success based on a variety of individual, team, and organizational factors.	X		X	X	X		
Quality Process Performance Model Development for F-16 Operational Flight Programs (p 81)	Bennett	Used multiple linear regression to predict the number of action items and escaped defects based on data related to the peer review process.	X	X	X	X	X	X	X
Modeling with TSP									
Uses of Monte Carlo Simulation for TSP teams (p 83)	Webb	Used Monte Carlo simulation to predict quality and schedule at both the individual and team level.	X	X		X	X		
Process Performance Models and PSP/TSP (p 84)	Smith	Used traditional TSP quality profiles and other in-process, phase-based data to predict schedule and final quality.		X	X				X
Process Performance Baselines and Models in PSP/TSP (p 87)	Akiyama	Used traditional TSP measures of quality including the product quality index to predict final schedule performance. Also showed modifications of using uncertainty distributions for some of the baselines to help in judging future performance.					X		X

2 Model Presentations

2.1 Discrete Event Simulation

Putting Process Performance Models to Work in the Real World: An Example and Recommendations

David M. Raffo, Portland State University

David Raffo's slides in this keynote presentation for the second workshop described his approach to process modeling and simulation, with an emphasis on the use of discrete event simulation to predict schedule variance and quality based on a variety of controllable factors. The complete set of healthy ingredients was evident during his presentation. The latter part of his presentation turned into a wide-ranging question and answer session with the rest of the workshop attendees, so much of what follows here covers his experiences with many other cases as well. These include more than two dozen process modeling projects in multiple organizations ranging from CMMI maturity levels 3 through 5.

Most of the organizations he discussed used process simulation in multiple projects. Some created small groups or departments to lead or perform that work. Their results ranged across a wide spectrum: some had great success integrating the models into their standard decision making processes, while others have achieved only limited use in a project or two. In some cases the results were never used. Success seemed to be positively correlated with how well organizations understood advanced methods and their willingness to make decisions on a factual basis.

In the primary example case, the organization was releasing defective products with high schedule variance in a CMMI maturity level 3 environment. Their defect-removal activities were based on unit test, where they faced considerable reliability problems. The project was in trouble. They needed to look at a variety of ways to reduce schedule variance and improve quality and had more than a dozen ideas to consider. In addition to their need to improve performance, they wanted a quantitative evaluation of the likelihood of success of their process improvement efforts.

A state-based discrete event model of large-scale commercial development processes was built to address that and other problems based on actual project data. It predicted project performance in terms of effort, task duration, and delivered defects. Part of a full business case analysis, it determined return on investment (ROI) and related financial performance of the proposed process changes. The model covered the full software development life cycle to a depth of two or more layers and around 30 process steps.

A process-level model was needed to capture complex relationships, multiple process paths, uncertainty, and the details of resources and artifacts. The tool used needed to provide a graphical depiction of the process while allowing for rapid yet easy model construction, modification, and support for sensitivity analysis. A model was built using Raffo's Process Analysis Tradeoff Tool (PATT). The model was created by a team that included the organization's own software engineers, in-house simulation experts, software engineering process group (SEPG) members, and both in-house and external consultants. In the end, the model was used by management, the SEPG, and the organization's software engineers.

The model helped the organization's personnel investigate the following questions:

- Will the process change improve project performance?
- What cost does the firm currently pay for conducting unit tests?
- Is partial implementation of the proposed process changes possible?
- How would potential learning curves affect the performance of the proposed process changes?
- Would alternative process changes offer greater improvement?
- Can the project benefit from reusing process artifacts?

Among other things, outcomes predicted by the models included the following:

- cost in staff-months of effort or full-time-equivalent staff used for development, inspections, testing, and rework
- numbers of defects by type across the life cycle
- delivered defects to the customer
- calendar months of project cycle time

Raffo also stressed the importance of addressing threats to data quality and integrity. For example, a detailed analysis of the data at the organization in his primary example showed a high degree of variation in the performance of different teams. The focus was kept on improvement by showing management average performance and spotlighting exemplary performance. However, the modelers had to adjust the team-level data because of that variability. In the process of doing so they found that the variation was driven by whether or not people had followed the existing process. That finding led to simulations predicting the effects of difference in process conformance and also helped to show that detailed unit test planning is not always necessary with a high performance team.

Much of the data needed for the example case and other cases on which Raffo has worked already existed in easily analyzable quantified format (e.g., in defect tracking databases and organizational accounting systems). However additional "data archeology" and "dumpster diving" often was needed to make use of hand-written forms and various unstructured documents such as inspection forms, process documents, project management tracking reports, and presentations. Raffo emphasized that ideas for appropriate x factors and data for parameter estimates in a model can come from many sources. Existing information within an organization's own current baseline is not the only source as data from the organization's own exemplary projects and pilot results can be particularly useful. Raffo is particularly fond of using expert judgments, especially to enable what-if sensitivity analyses in important areas where no data typically are collected. If proper care is taken, industry data sets from comparable organizations can be used as well.

The pluses and minuses of different modeling techniques were thoroughly discussed during the presentation. For example, process simulation models can rely on state-based, discrete-event, or system-dynamics techniques, or contain aspects of all three. They can integrate various submodels that are used to predict specific aspects of process performance and product quality outcomes. However, simple Monte Carlo or spreadsheet what-if analyses can be used depending on the needs of the particular organization and situation. Many different kinds of mathematical models

can be used, including queuing networks using Markov chains; forecasting models using moving average or weighted moving averages; time series models accounting for autocorrelation; regression models of independent factors; or structural models based on theory.

Process simulation models can be quite useful in many areas, particularly in optimizing verification and validation strategy and process improvement investments, providing quantitative justification for change, and minimizing risk. They can be used for fine tuning at the subprocess level as well as for decision support and tradeoff analysis. The initial cost of model development can range from a month or two of staff effort to a year depending on the scope of the modeling effort. Tools can range from $5,000 to $50,000 depending on the level of capability provided. At the same time, model results can and have saved organizations millions of dollars through resultant improvements in processes and return on investment.

Yet barriers often must be overcome to encourage the creation and productive use of process simulation models. Important stakeholders may not have confidence in a model, especially if they do not like the model results or their implications. The data on which the models rely may be viewed as imperfect and their results not credible, especially when the model results appear to be unintuitive. And such skepticism can be aggravated with engineers who do not like the prospect of additional process steps affecting their work. There also are misconceptions of just what is necessary along with perceived risks about untested technology and the credibility of predictive models. Perceived cost is often a concern as well.

So what can be done? Raffo emphasized the importance of building trust in the models and their results. Key stakeholders must be included from the beginning in any modeling effort to build consensus as well as credibility. That includes in particular management team members from all levels concerned with project-level performance. Equally important, understandable models must be created with reasonable assumptions and generally accepted data for managers and engineers who are not modeling experts but whose own work may be affected. Clear graphical depictions, straightforward model construction and modification, and support for rapid assessment of alternatives can be crucial. Fortunately, once they are built, model tailoring often requires much less expertise than what is needed for building them, and several commercial tools now exist that can facilitate building simulation models much more rapidly.

The purpose for building a model determines the degrees of precision and accuracy needed. The amount of certainty required for tracking projects or creating estimates to bid on contracts is much higher than selecting among process improvement alternatives, so the initial cost is not necessarily high. Regardless, modelers must stay focused on the critical needs of their organizations and customers. Model development should proceed iteratively so beneficial results can be delivered early and often.

Putting Process Performance Models to Work: A NASA Example

David M. Raffo, Portland State University

In this presentation, David Raffo shared his experiences in using discrete event simulation to predict the cost, schedule, and quality impacts of adopting the QuARs requirements analysis tool. His presentation covered all the healthy ingredients for PPMs.

Raffo described a collaboration with a NASA site concerned about how best to improve the product quality and business value of its verification and validation (V & V) processes. Raffo worked with NASA staff and the SEI on a project using his PATT modeling tool to help NASA optimize its V & V strategy by predicting the likely outcomes of decisions about technology adoption and tradeoffs among process improvement alternatives. NASA needed to decide whether or not to adopt a requirements analysis tool called the Quality Analyser for Requirements Specification (QuARS) and if so how best to use it. Detailed project quality and performance goals included improved quality assurance, increased coverage, and reduced risk. A pilot study was done to help answers these questions:

- Would the return on investment justify the cost of the tool and associated training and process change?

- If so, what project conditions and process would offer the best results?

QuARS, developed at the Italian National Research Council, is an automated tool that performs lexical and syntactical analyses to identify defects in requirements and specifications documents. QuARS searches for several kinds of potential defects that display properties such as ambiguity, understandability, and insufficiently detailed specifications. The tool is interactive and allows the user to revise offending statements during the analysis session. Detection and correction of defects with QuARS can be made both during the realization of a requirements document and during the review or inspection of the document.

Stakeholders who were the intended users of the model and audience of its outcomes included NASA IV & V, headquarters program and project management, IV & V contractor organizations, and NASA project contractor organizations. A process-level model that captured complex relationships, multiple process paths, details of resources and artifacts, and uncertainty was needed to address NASA's goals. Given NASA's resource constraints and the need to work with practitioners whose forte was not in modeling and simulation, there also was a need for a clear graphical depiction of the development process along with ease-in construction, easy modification of the model, and support for sensitivity analysis.

As noted in Raffo's previous presentation, the PATT architecture is well suited for rapid model creation and maintainability. It also supports the integration of various models and data that may be of differing quality. PATT was used to build a model of a large-scale software development process based on the IEEE 12207 process standard with additional detail beyond the standard for IV & V. Since 12207 was already implemented as an optional template in the PATT product suite, adding an IV & V layer on top of it was much less costly than it would have been to implement the entire model from scratch. A visual representation of the model is shown in Figure 2.
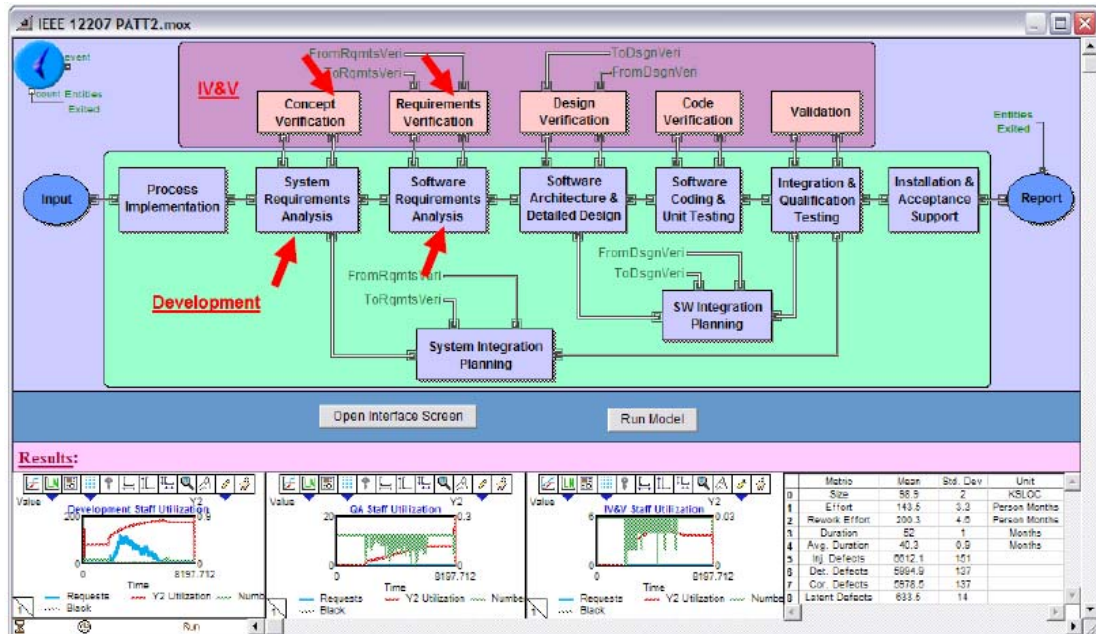
Figure 2: NASA Model with IV & V Layer on IEEE 12207 Life-cycle Process

The model elements can be entered graphically, and further details can be specified formulaically if necessary. Simulation results are summarized in the lower portion of the figure. The model has a depth of over three layers with about 86 process steps. The data were drawn from NASA IV & V data (initially for five projects); NASA high-level project data down to the CSCI and CSC level; and lower level baseline software development process data from published industry data sets. Project size was based on varying project characteristics, and project complexity was based on expert judgment. Project performance was predicted in terms of development effort, task duration, and delivered defects.

Outcomes predicted by the process performance model included the following:

- cost expressed in person months of effort and full time equivalent staff used for development, inspections, testing, rework, and IV & V
- number of defects by type across the life cycle as well as delivered defects to the customer
- calendar months of project cycle time

X factors included the following:

- development and rework effort allocated by phase across the software development life cycle
- IV & V issue resolution (i.e., open, in process, or resolved)
- project duration (estimated based on industry data)
- IV & V issues tracked
- where defects arose and were found (subject to data availability)
- total number of defects injected, detected, and removed for the project (estimated based on industry data)
- defects distinguished by severity (estimated based on NASA percentage data)

Simulations were run applying QuARS as a V & V activity within the project and as an IV & V activity outside of the project. In both instances simulations were run using QuARS at both the systems and software requirements phases, in each case assuming either a 100% or a 50% requirements inspection and use either before or after inspection. Case analyses were done for injection of the types of defects that QuARS is capable of finding. The analyses were straightforward and typically took about one to two weeks. The results were sensitive to the percentage of project inspected, the amount of defects that had been injected before analysis with QuARS, labor rates, rework costs, and the hurdle rate used by NASA. All of these are at least potentially controllable x factors.

The potential benefits of this work were substantial for NASA. It typically takes three to twelve person months to create a model of this complexity. About one person month was needed for each case analysis at NASA; however use of the model identified a sweet spot in project development that could save the agency $1M for similar projects. The optimized process improvements more than doubled the expected ROI. The value to the project at a 20% hurdle rate (the minimum acceptable rate of return) ranged from $280K to $930K in V & V mode and $266K to $540K in IV & V mode. Application of QuARS during systems and software requirements offered the best value at NASA, where the sweet spot was to apply QuARS after software requirements in V & V rather than IV & V. Using QuARS before requirements inspection rather than after could yield a savings of approximately 10% to 15%, with about a 3% increased performance for projects without IV & V. With a 50% requirements inspection rate, there could be approximately a 20% to 30% increased savings in effort, with a 17% to 42% reduction in latent defects.

Moreover, the savings could be even greater if more defects were injected prior to using QuARS. The estimated savings in effort range between 28% and 36% and quality savings range between 28% and 38% when applied for V & V. The effort savings were estimated to be greater for IV & V, ranging between 35% and 43%, although quality savings were estimated to range between 26% and 36%.

What should NASA be willing to pay for QuARS? While the cost of QuARS remained unset at the time of the work with NASA, up to $266K likely would be a clear win, which is the low end of the estimated savings if the requirements analysis was done at IV & V with a hurdle rate of 20%.

QuARS was being integrated into an existing IV & V planning process at the time this work was done, and some contractors and IV & V managers stood to gain or lose resources. A new director then changed the business model for IV & V when faced with significant research funding cuts. That ended further work on this project, at least in the short run. Management changes and redirection changed things; however, the situation may change again thanks to a continuing engagement in process work at NASA and a reemphasis there on ROI, process improvement, and optimization.

A Process Performance Model Used Within Lockheed Martin MS2

Stephen Austin and Bob Beckley, Lockheed Martin

This presentation by Stephen Austin discussed Lockheed Martin's desire to shorten the time it took to release engineering drawings by using a discrete event simulation of the drawing control and release process. All of the healthy ingredients of a PPM were evident in the presentation. A value stream map structured improvement activity (SIA) was conducted on the current drawing release process, and the SIA resulted in an improved future state. Due to the number of variables

in the process, previous manual attempts to predict average cycle time, touch time, touch labor, and throughput were unsuccessful. The Drawing Release Process Model was developed to make predictions about the future state of the drawing release process for the following:

- average cycle time per drawing
- average touch time per drawing
- average touch labor per drawing
- average weekly throughput for the process (i.e., drawings completing the process per week)

It also provided insight into the most likely task bottlenecks and resource contention issues. The inputs (or drivers) for this model were the resources and touch time used for each detailed step in the process.

The Drawing Release Process Model provided predictions and performance measures for the outcomes discussed above. The predictions were significantly lower than the objectives set by management. For example, the throughput rate was 7.56 drawings per week as compared to the objective of 100 per week. It also predicted that an initial task (i.e., the drawing review) was bottlenecking the entire process because of severe resource limitation (a single reproducibility engineer was available) and average task effort (it took four hours to review a drawing). Other major process risk points were the resource requirements needed to

- draft the drawing
- make major or minor changes to the drawing
- conduct the drawing release review

Figure 3 shows a model representing the initial state.
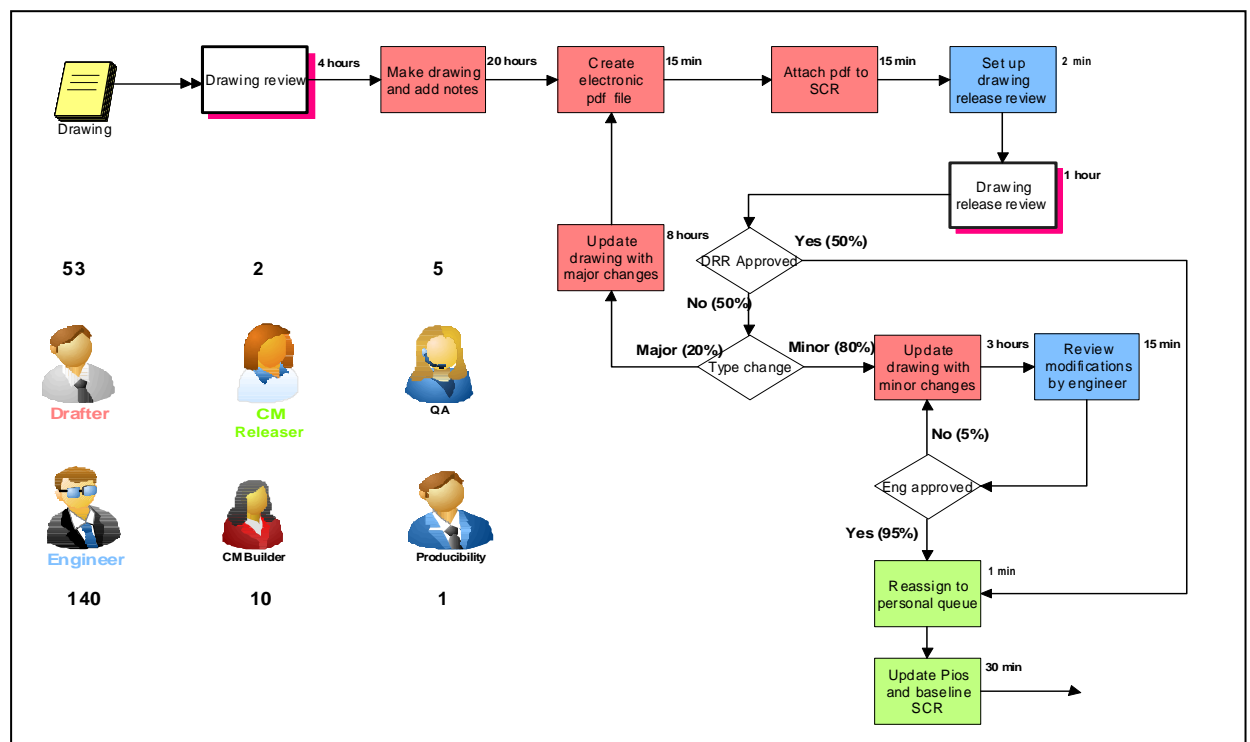
Figure 3:   Drawing Release Process, Initial State

Once the model had identified the bottlenecks, changes were made to the initial task so that a more available resource was used to accomplish the same task, and the throughput improved from 7.56 to 64 drawings per week. The model was then used to determine the minimum set of resources necessary to achieve a throughput of 100 drawings per week.

Figure 4 shows the process after modeling, along with a comparison of throughput improvements. The process flow remained the same, but resources were re-allocated to reduce the bottlenecks in the process. The before and after throughputs show a dramatic change. The optimized solution is also shown, although it is not practical to implement because it would require too many drafters.
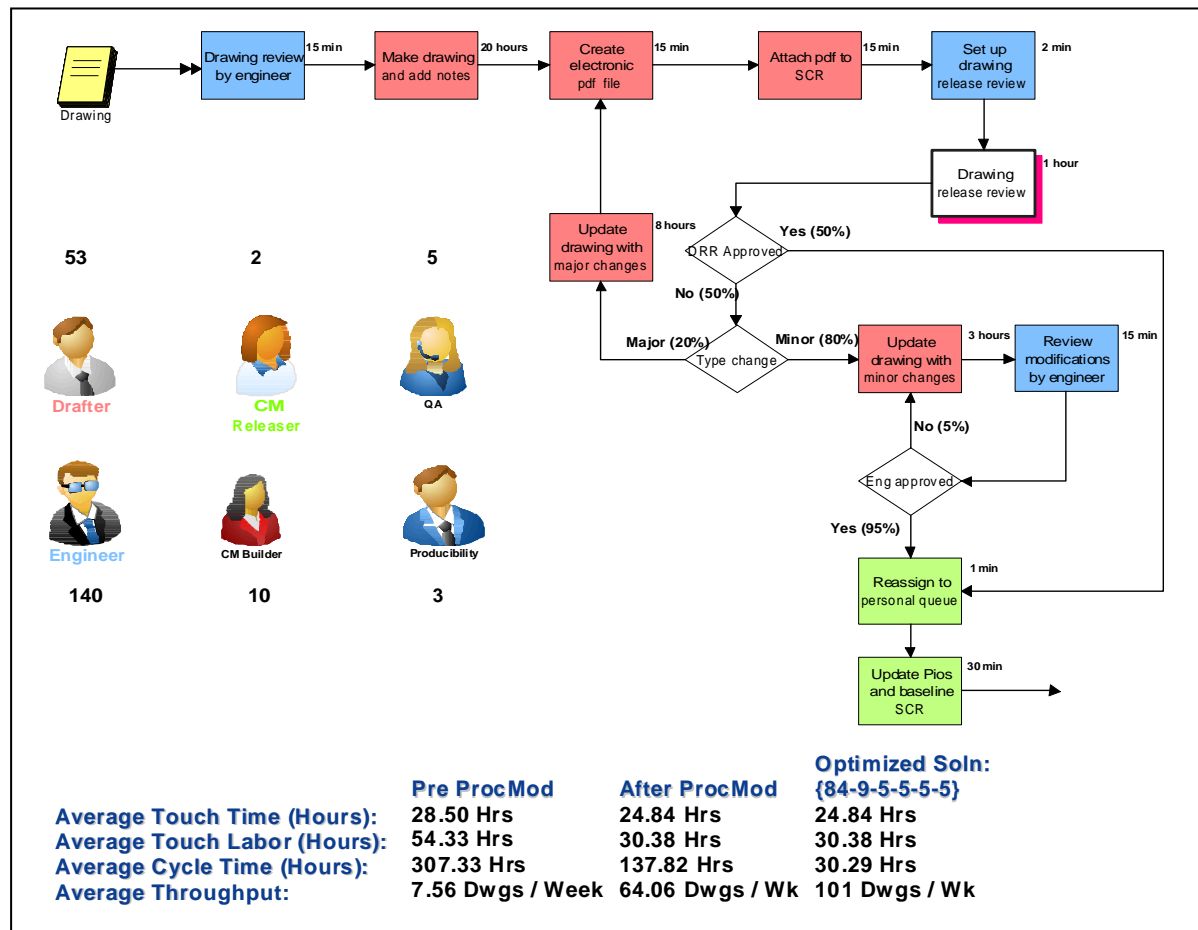
Figure 4:   Drawing Release Process After Process Modeling

The model provides averages, standard deviations, minimums, maximums, skewness, and high and low 95% confidence intervals. Sample data for 999 iterations of a 10-week cycle for a specific set of resources are shown in Table 2.

Table 2:   Statistical Results from a Simulation Run of the Drawing Release Model

```
                       Average    Average  Average
                        Cycle      Touch    Touch
     Entity      Qty     Time       Time    Labor
     Name   Processed  (Minutes)  (Minutes) (hours)
     -------  ---------  ---------  ---------  -------
     Drawing   206.84    10251.2    1414.82    29.76  (Average)
     Drawing     5.51     199.37      13.50     0.27  (Std. Dev.)
     Drawing      184     9723.39    1374.31    28.88  (Min)
     Drawing      224    11100.9     1461.10    30.81  (Max)
     Drawing    -0.13        0.36       0.14     0.21  (Skewness)
     Drawing   206.48    10238.3     1413.95    29.74  (95% C.I. Low)
     Drawing    207.2    10264.1     1415.69    29.78  (95% C.I. High)
```

This model has since been updated and measurements have been generated to make cycle time predictions between many different points in the process, which are also action log points in the configuration management system. Other development programs can run this model with adjustments made to reflect their resources to see what drawing release rate could be achieved, which helps them with scheduling. Programs can also run it periodically to assess the risk of meeting drawing release throughput requirements.

Austin shared some lessons learned during the adoption experience. The SIA consisted of a week-long Kaizen event.[2] Less than 10 staff hours were required to create the simple model shown in this presentation because the Lockheed Martin team had ready access to an expert already familiar with discrete event simulation, and, more specifically, the ProcessModel tool. Another lesson learned came from the challenge of getting high fidelity data to populate the discrete event simulation model. The pursuit of this model reinforced the need to emphasize a rich, high-quality measurement program consisting of the measures most likely needed for such modeling. Finally, it was recognized that increased communication to management about the business value of such modeling was needed. Management often makes decisions about where to invest the limited process improvement dollars and, as such, they need to better understand when and where to deploy this type of modeling.

2.2   Monte Carlo

Compose Process Model

Shankar Mallapur, Johnson Varghese, and Gayathri Pallail, Accenture Private Services Limited, India

In this presentation, the authors showed how Monte Carlo simulation was used to predict defects and cycle time based on the controllable factors of development and peer review. All of the healthy ingredients for PPMs were included in the model. A process performance model, called

---

[2]   Now widely used in Lean Six Sigma, Kaizen was first implemented in Japan after World War II. Kaizen focuses on making changes incrementally, monitoring results, and quickly adjusting activities accordingly [2].

the Compose Process Model, was used to measure progress toward the high priority objectives set for the organization. These quality and process performance objectives included the following:

- improving the quality of the work delivered to clients by
    - reducing the defects injected in the life-cycle phases of the project
    - improving review and testing effectiveness and efficiency
- on-time delivery
- cost-effective delivery

The model predicted a range of performance results, taking into account project-specific priorities for quality, cycle time, and effort. Additionally, tradeoffs between competing priorities were handled by the model. Based on model results, focused improvement initiatives were implemented to help meet the quality and process performance objectives established by the organization. The model was essential in helping the organization understand how closely these initiatives could deliver on expected results.

The Compose Process Model is a composite model that integrates multiple models for defect prediction and cycle time prediction. It is integrated with Crystal Ball and Opt Quest, allowing the project management team to estimate the probability of meeting project objectives using Monte Carlo simulation.

The first submodel within the Compose Process Model, called the Defect Prediction Model, predicted the expected range of defect injection and defect detection values at different stages of the life cycle of the project. This helped predict the progress toward improving the quality of work. The major factors of the Defect Prediction Model are technology, phase effort, review technique, % review effort, and % testing effort. This model is a combination of phase-level defect-prediction models. Table 3 lists the phase-level defect prediction models, process or subprocesses modeled, statistical techniques used to build the model, and the operational purpose of the model. The Defect Prediction Model was designed for use by the project management team during the planning phase and after completion of each of the subprocesses listed in Figure 4.

Table 3:   Defect Prediction Model

Phase Level Defect Detection Model Name	Process Modeled	Statistical Technique used to Build the Model	Operational Purpose of the Model
Analyze Defect Injection Prediction Model	Analyze Application	Correlation & Regression	Predicts the expected range of values of defect injection at the analyze phase of the project
Analyze Defect Detection Prediction Model	Analyze Application – Peer Review	Correlation & Regression	Predicts the expected range of values of defect detection at the analyze phase of the project
Design Defect Injection Prediction Model	Design Application	Correlation & Regression	Predicts the expected range of values of defect injection at the design phase of the project
Design Defect Detection Prediction Model	Design Application – Peer Review	Correlation & Regression & ANOVA	Predicts the expected range of values of defect detection at the design phase of the project
Build Defect Injection Prediction Model	Build Application	Correlation & Regression	Predicts the expected range of values of defect injection at the build phase of the project
Build Defect Detection Prediction Model	Build Application – Peer Review	Correlation-Regression & ANOVA	Predicts the expected range of values of defect detection at the build phase of the project
Component/Unit Test Defect Detection Prediction Model	Prepare & Execute Component Test	Correlation & Regression	Predicts the expected range of values of defect detection at the Component/Unit Test phase of the project
Assembly/Integration Test Defect Detection Prediction Model	Prepare & Execute Assembly Test	Correlation & Regression	Predicts the expected range of values of defect detection at the Assembly/Integration Test phase of the project
Product/System Test Defect Detection Prediction Model	Prepare & Execute Product Test	Correlation & Regression	Predicts the expected range of values of defect detection at the Product/System Test phase of the project

In the phase-level defect detection models, the phase effort was computed based on the productivity of a particular phase as shown in the organization's process performance baselines. Defect injection in a particular phase was estimated based on the phase effort. For the design and build phases, defect injection was predicted based on the phase effort and technology used.

Defect detection was estimated based on the review effectiveness for the phase, while review effectiveness was predicted based on the % review effort and the review type planned. Defect detection was then computed as the multiplied product of defects injected and review effectiveness.

The defect leakage from each phase was distributed to the future phases based on the review effectiveness. Residual defects were distributed across the testing phases based on the testing effectiveness, and testing effectiveness was computed based on the testing effort planned.

The Cycle Time Prediction Model predicted the cycle time required for a phase or a subprocess based on the phase effort and the team size. This model was used by the project management team during the planning phase and after completion of every subprocess to predict the progress toward quality and on-time delivery with optimum productivity. It was also designed for use by the project management team during the planning phase and after completion of each of the subprocesses listed in Table 4.

Table 4:   Cycle Time Prediction Model

Model Name	Process Modeled	Statistical Technique used to build the model	Operational Purpose of the model
Cycletime Prediction Model	Lifecycle Process from Analyze Application till Prepare & Execute Product Test	Correlation & Regression	Predicts the expected range of values of cycletime at various life cycle phases. Prediction is done based on phase effort and team size.

The Compose Process Model was built with Crystal Ball to run the optimization in conjunction with Monte Carlo simulation. The model provides built-in, multiple options for design and build review processes and other subprocesses. Crystal Ball simulation provides estimates of the probability that the project will meet the project objectives. Sensitivity analysis in Crystal Ball helps to identify the most critical subprocesses and the candidate critical controllable factors for the selected subprocesses. The performance of these subprocesses are then statistically monitored to ensure that the process is capable of delivering the project objectives. Optimization using OptQuest is used to make decisions when composing the project's defined process to optimally meet the project's objectives. Actual process performance data is fed back into the model after every subprocess execution. Optimization and simulation are periodically re-run to ensure that the relevance and capability of the project's defined process will meet the project's objectives.

If the model's probability outcome is not within the acceptable range, the management team should take the steps listed below.

1. Conduct a risk analysis to understand the impact if the probability of being outside the acceptable range is low.
2. Conduct a root cause analysis and implement preventive actions if the probability of being outside the acceptable range is higher.
3. If the probability is far higher,
   a. implement innovations to improve the probability of meeting the project objectives
   b. renegotiate the objectives with the client

The Compose Process Model provides the following benefits to projects:

• ensures the right use of data to make the right decision at the right time
• helps set SMART goals for subprocess performance based on the project's objectives
• enables a proactive approach to reducing cost and quality risks in the project
• helps in negotiation of project objectives with clients
• alerts the organization to the need for re-negotiation with the client or the business as soon as the process performance deviates from the expected range
• helps prioritize organization- and project-level goals, followed by the selection of the correct subprocesses for statistical monitoring
• provides a high degree of knowledge of the subprocesses that will produce high defect injection possibilities during the planning stage
• helps reduce rework through a proactive analysis of the models

- allows the organization to reduce the impact of a small deviation in process performance by resolving it before it cascades into future life-cycle phases
- provides insight into how effective planned improvement initiatives will be

From an adoption experience standpoint, the Compose Process Model took approximately two staff months of labor to develop and required approximately six hours of effort to collect data from the involved stakeholders each time the model was used to make a prediction. The primary challenge was teaching project managers to use the model, conduct predictions, analyze results, and determine the proper actions to take.

Scheduling Analysis of Variability Engine (SAVE)

Neal Mackertich and Michael Campo, Raytheon Integrated Defense Systems (IDS)

Neal Mackertich presented a Monte Carlo simulation model developed at IDS that helps projects predict schedule outcomes based on the uncertainty of individual work tasks in the program. The presentation did not directly share evidence of controllable factors or the conduct of what-if analysis, but subsequent discussions indicated that the model would be evolved to include those two healthy ingredients of PPMs.

The model, called the Scheduling Analysis of Variability Engine (SAVE) model, supported Raytheon's goal of developing increasingly complex systems with smaller performance margins that are also open and adaptable and meet the users' requirements in the shortest time with the highest reliability and at the lowest cost.

Of these challenges, the most daunting was schedule pressure. Countless projects fall under the category of "yes we can do it, but not with that schedule." Traditionally, schedules in Raytheon have been managed deterministically by the task manager, limiting the ability of the organization to assess the risk and opportunity involved, perform sensitivity analysis, and implement strategies for risk mitigation and opportunity capture. Using an enterprise-wide license for Crystal Ball and an internally developed algorithm and interface, the Raytheon SAVE model enables projects to

- statistically predict their likelihood of meeting schedule milestones
- identify task drivers based on their contribution to overall cycle time and percentage of time spent on the critical path
- develop strategies for mitigating the identified risk based on a captured set of best practices from the SAVE model

Inputs to the SAVE model included the statistical estimation of individual task activity duration and the overall scheduling requirement. Individual task durations were expressed as probability distributions. Historical data and engineering experience both suggest use of an underlying log-normal distribution, but estimation of the standard deviation at the individual task level is a barrier.

Individual task sequence relationships were used for the overall scheduling requirement. An understanding of the overall flow and sequencing of the product development activities associated with the achievement of project delivery milestones forms the basis for the overall scheduling

model. The SAVE model algorithm designates each individual task activity with an identification number used when representing its precedence.

Expert estimates were used when historical data were not available. The triangular distribution is the most appropriate probability distribution for the expert estimation. It provides a better fit for current integrated deployment efforts since stakeholders are generally more comfortable with providing estimates of the shortest, most likely, and longest time durations. Long term, data collected from the deployment efforts will be analyzed to understand how the estimation may be improved.

The primary output of the model was the prediction interval estimate of schedule performance (generated from Monte Carlo simulation) using individual task duration probability estimation and an understanding of the individual task sequence relationships. Mackertich demonstrated the developed model at the workshop and shared quantitative and qualitative results from its organizational deployment. A screenshot of the SAVE model interface is shown in Figure 5.
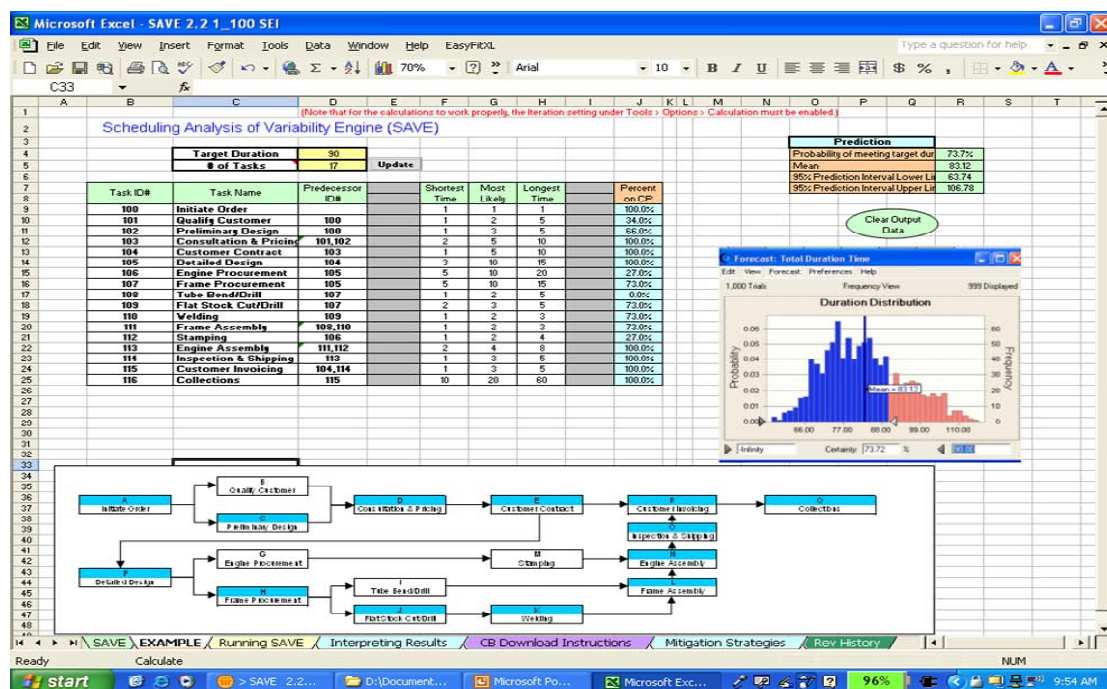
Figure 5:   SAVE Model Interface

Crystal Ball was selected for simulation modeling because of its availability (Raytheon has a corporate license) and general ease of use for project teams. The SAVE model transfer function was developed for each instance using inputted project task predecessor information. Monte Carlo simulation runs were then generated using this transfer function and individual task duration probability distributions. The SAVE model includes the following components:

- Crystal Ball download instructions
- step-by-step guidance for projects on how to run SAVE
- guidance for interpreting the results
- a list of mitigation strategies

SAVE is used by integrated project teams made up of systems, software, hardware, and quality engineering staff during project planning and execution in order to predict, manage, and mitigate risks in achieving project schedule milestones. Stakeholder groups and projects found the model both easy to use and conceptually aligned with project issues. As a direct result of SAVE and critical chain modeling and analysis, projects identified and implemented specific improvements to their processes and execution that enabled cycle time reductions in the following areas:

- process redesign based on critical path dependency
- resource reallocation and conflict resolution
- increased investment upfront in planning, analysis, and training activities in order to enable execution speed ("festina lente")
- minimizing churn through enhanced peer review

In addition to its use with projects, SAVE was used upfront during the bid and proposal phase and was also used by engineering management during schedule negotiations with program management. Significant qualitative benefits from SAVE integrated deployment should not be underestimated: project leads and teams began thinking and behaving differently with respect to their analysis of schedule risk and opportunity. Engineering process funding was invested in the development and deployment of the SAVE model and critical chain project management, resulting in a 15 – 40% reduction in cycle time duration against baseline. Additional process funding is now set aside for maintaining, updating, and supporting the continued project deployment of SAVE.

The challenges listed below were noted in developing and deploying the SAVE model.

- The organization had to learn to think differently about risk and opportunity analysis.
- Project teams had to create the pull for the model to be used.
- Boundary-less thinking was required to enable the interdependent execution.
- Estimating the input factor probability distribution was sometimes difficult.
- Model usage was dependent on proper data collection, quality, and integrity.
- Crystal Ball licensing and download issues had to be overcome.
- Discipline was needed to document evidence of the use of the model.
- Quantitative risk and benefits analysis had to be understood and acted on correctly.

The Raytheon team that deployed the SAVE Model also noted what worked well, including

- upfront leadership engagement and commitment
- a focus on business value
- a multi-skilled and cross-functional model development team
- use of the Raytheon IDS engineering CMMI OID process area implementation
- upfront project team involvement
- the simplicity of the model interface
- the Crystal Ball user-friendly interface
- communication, communication, communication

The effective development and deployment of process performance models sparked increased interest in the following:

- interdependent, integrated business execution
- statistically based project performance risk assessment
- identification of leading indicators that statistically drive project performance
- statistical modeling and analysis supporting tools and training
- follow-on statistical modeling and analysis efforts
- business investment in process performance modeling throughout the product development life cycle.

Process Performance Models: Process, Results, and Lessons Learned with the System Lifecycle Analysis Model (SLAM)

Neal Mackertich, Michael Campo, and Rachel Beitz, Raytheon Integrated Defense Systems (IDS)

This presentation showed how a multiple linear regression model combined with Monte Carlo simulation can predict final cost and schedule performance of programs based on requirements volatility and the degree of overlap of the requirements and design phases. Such a model can be used to identify the risk of proceeding with development prematurely. The System Lifecycle Analysis Model (SLAM) was developed and deployed at Raytheon Integrated Defense Systems to quantitatively assess the cost-performance risks associated with requirements volatility and the life-cycle overlap of requirements and design. Neal Mackertich discussed the use of SLAM by integrated project teams made up of members from systems, software, hardware, and quality engineering during project planning and execution to predict, manage, and mitigate risk in achieving project cost objectives. The engineering process group used SLAM to estimate benefits of process improvement proposals and to measure changes in performance due to process improvements.

Two important factors influencing the ability of IDS programs to meet cost performance index (CPI) and software process improvement (SPI) business and project cost and schedule performance goals are

1. aggressive program schedules with increased overlap between life-cycle phases. Design typically begins before requirements are complete.
2. volatility of requirements, which causes rework for software and hardware development

Projects have been unable to quantify the cost performance risk associated with these two conditions. SLAM was developed to help projects quantitatively identify and assess this risk so it can be mitigated.

The outcome prediction from this model is a confidence interval estimation of cost performance generated using Monte Carlo simulation with a developed multi-factor regression model. Key stakeholder audiences for the model are systems, software, hardware, and quality engineering teams.

Factors used in the model are requirements volatility and requirements/design life-cycle overlap. Requirements volatility is a required measure collected and reported by every development project across Raytheon. During piloting, requirements volatility data was collected at the confi-

guration item or module level. Requirements/design life-cycle overlap is a non-standard project measurement collected and analyzed for the first time during the SLAM development piloting and deployment. Essentially, the overlap represents risk and is quantified by determining the amount of relative project effort charged during the overlap period compared to the total project planned effort. Obviously, the greater the percentage of total project effort expended during the overlap, the greater the risk of rework and other negative consequences to the project.

A mathematical function of the input factors was reasonably well correlated with the output responses using linear regression techniques, with an adjusted r-squared value equal to 0.65. Additionally, collected project data from SLAM piloting and deployment further confirmed the strength of the underlying statistical relationship.

Crystal Ball was used as the modeling tool. A Monte Carlo simulation model was developed with an Excel-based user interface using the correlated regression equation and estimates of mean and variance for each of the factors (from the collected data).

Model inputs included

- estimated % design complete at systems requirements release, with confidence range of +/-5%, 10%, or 15%
- requirements volatility estimate (i.e., the best estimate based on historical baseline; variance estimates built into model based on historical actuals)
- projected software/hardware cost performance (mean, standard deviation) with 95% upper and lower confidence interval limits

Model output consisted of projected software or hardware cost performance (mean, standard deviation) with 95% upper and lower confidence interval limits.

Based on the SLAM results, mitigation strategies were developed for reducing requirements volatility and the degree of overlap. Products might release slightly later, but with exceptional quality (i.e., low volatility) resulting in reduced churn and rework. All of the Raytheon IDS stakeholder groups and projects found SLAM easy to use and conceptually aligned with the project issues. Use of this model led Raytheon IDS to conclude that process improvement efforts to reduce requirements volatility would provide quantifiably significant program performance improvement.

As a direct result of SLAM modeling and analysis, a majority of projects identified and implemented specific improvements to the projects' defined processes (which subsequently reduced requirements volatility), including the following:

- improvement of the systems requirements review process and execution
- increased degree of concurrent engineering between systems and software/hardware engineering
- increased use of critical chain concepts in order to shorten design cycle time
- consistent use of SLAM by all projects in performing risk assessments

Mackertich reported that with one exception (related to data integrity issues), all projects found model predictions to be aligned with project actuals. As a result of the initial success, Raytheon

IDS plans to fine-tune this modeling approach by further stratifying data based on project types and environments, thereby creating a more specialized model for each stratification.

In addition to deployed projects, SLAM data was cited by the IDS software engineering director during schedule negotiations with program management. As such, significant qualitative benefits from its integrated engineering deployment should not be underestimated.

Evaluation of SPI Efforts in Small & Medium Organizations

Pedro E. Colla, Instituto Universitario Aeronáutico and Universidad Tecnológica Nacional-Facultad Regional Santa Fé; and Jorge Marcelo Montagna, INGAR-Instituto de Desarrollo y Diseño, Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional-Facultad Regional Santa Fé

In this presentation, Pedro Colla discussed his use of both a deterministic model based on mathematical transfer functions and a stochastic (Monte Carlo) scenario model. These models were used to predict the likelihood of net present value from the adoption of CMMI-based process improvement initiatives in small and medium software development organizations in Argentina. The nature of this modeling meant that the following healthy ingredients were not implemented: connect upstream activity with downstream activity and enable projects to achieve mid-course corrections to ensure project success. Rather than focusing on a single organization, the work centered on helping the broader community evaluate the likely outcome of process improvement efforts. The intended audience for the work included executives from small and medium enterprises and the SPI practitioners who support those doing business case evaluations. Since small and medium enterprises often have limited resources, another audience for the work was policy makers at the industry level who could pool their resources in identifying shared drivers and data collection. Government officials who were evaluating policies aimed at fostering wider adoption of process improvement efforts in Argentina were also part of the audience. Cooperation is necessary since most software organizations in Argentina are quite small, with a median of 20 people per organization.

Like many small and medium enterprises, those in Argentina have faced a leadership dilemma in justifying the investment required for process improvement. There is continuing tension between short-term survival and long-term competitiveness. While the benefits of process improvement often are understood, organizations remain reluctant to embrace improvement initiatives. Common concerns include the amount of time and investment needed. The affordability and suitability of formal frameworks (including CMMI) at small and medium enterprises is often questioned. Organizations often focus on product rather than process, and so far they have lacked credible evidence about the relationship between the two. They also lacked the tools and infrastructure they would need to assess and mitigate risk in organizations such as their own.

The modeling described by Colla initially focused on the effects of process improvement in Argentina on the following hypothesized results:

- increased income from new customers or additional work from existing customers
- operational efficiency as measured by fewer defects and shorter cycle time, modeled as productivity
- value protection and reduced delivery risk

- intangibles, (e.g., brand build-up, employee morale, and organizational maturity)

A large part of the effort was devoted to defining a deterministic dynamic model based on mathematical transfer functions. The intent was to integrate the analytical work into a single framework. Stochastic functions were introduced through Monte Carlo analysis. Variation of several model parameters was addressed, using actual distributions when data were available and triangular distributions when they were not. Sensitivity analyses were performed on all major x factors. The main outcome factor was the probability of a positive net present value (NPV > 0). An iterative approach was used to cover all of the major x factors. Tools used included VenSim[3] for building the deterministic model, SimulAr[4] for model testing and debugging, and GoldSim[5] for the stochastic evaluation.

Return on investment (ROI), often presented as a simple conceptual split between good and bad, is frequently cited as the major criterion to evaluate investment. However ROI alone has limited value, especially for small and medium enterprises where time and risk factors such as investment horizons, limited financial resources, and other dependencies are critical. Net present value properly factors in time and risk premiums for money. It thus captures an organization's situation with respect to long-term survival factors, short-term sources of risk, and other intangibles. NPV > 0 also provided a good distinction between good and bad for the modeling that Colla described.

X factors in the model included costs and effort of implementing and sustaining the process improvement initiatives; engineering process group activities; preparation and delivery of training; and appraisal costs. Other x factors mentioned in the presentation were cost per engineer, organization size, and CMMI maturity level, along with three additional financial measures (opportunity cost, risk-free rate, and investment horizon). All of these, with the exception of appraisal cost and opportunity cost, were treated as controllable factors. Interestingly enough, the implementation costs were not related to organization size. With the exception of effort to implement, there were no significant differences between low and high maturity level initiatives. All of these x factors were based on historical or survey data. With the exception just noted, all were found to be statistically independent.

However, maturity level also was correlated with several other x factors that were treated as interim "little y" factors. In addition to effort to implement the improvement, these include improvement initiative implementation time, recovery time, a risk variation factor, and the likelihood of achieving an appraised maturity level. Efficiency gains in productivity were expressed relative to maturity level. All of these were treated as controllable with the exception of maturity level and recovery time, which is a functionally derived measure. Historical or survey data were available for all of these measures with the exception of the risk variation factor.

Monte Carlo results from a typical run predicted NPV as a function of time. Probability distributions of different NPV outcomes were examined, with an emphasis on the probability of NPV > 0. NPV was sensitive to differences in organizational factors, particularly organization size, investment horizon, and growth. For example, as seen in Figure 6, the model predicts about a 35% probability that a positive NPV will be achieved in about 36 months for a maturity level 2 organi-

---

[3] Further information about VenSim is available at http://www.vensim.com/.

[4] Further information about SimulAr is available at http://www.simularsoft.com.ar/SimulAr1e.htm.

[5] Further information about GoldSim is available at http://www.goldsim.com/.

zation with fewer than 50 people. Secondary effects were predicted based on current opportunity cost and cost per engineer.

Probability of NPV>0
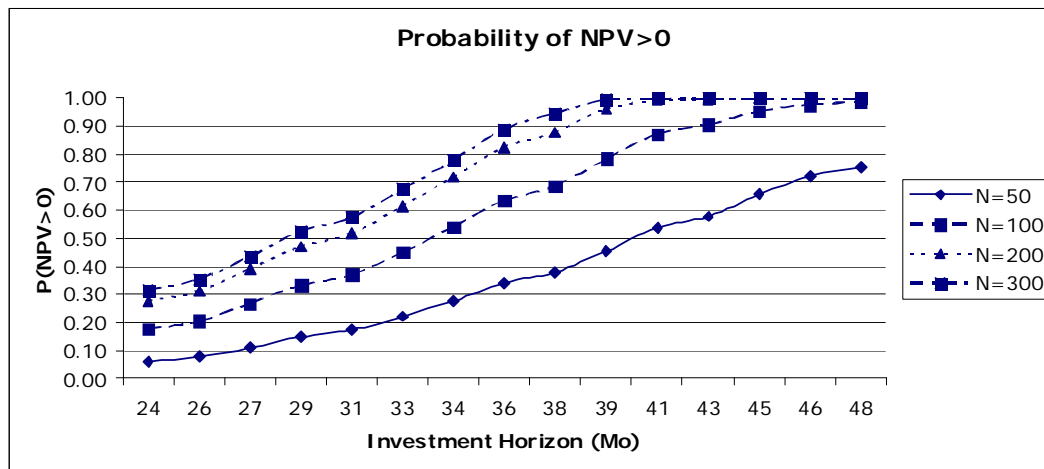
Figure 6: Organization Size and Investment Horizon

Larger organizations are more likely to get NPV > 0 faster. This is not surprising since larger organizations have more resources to devote to faster payback on their investments. Interestingly, the simulations predict that initial and higher maturity organizations will require similar effort and cycle time to achieve positive NPV, which is consistent with Argentine survey data.

Colla also discussed the limitations of the modeling that had been done at the time of the presentation. The work was based largely on historical data augmented by survey data collected expressly for the modeling effort. A substantial amount of information was available for parameters such as implementation time; however, much less was available for others such as the training parameters. Moreover very few data points were available for some of the parameters, contradictory data existed for estimating some of the transfer functions, and insufficient data were available to calibrate the model to address finer grained questions (e.g., about the possible effects of partial CMMI-based implementations by process area). Additional verification and validation clearly remained necessary.

That said, several useful conclusions were warranted. While formulation of the models remains complex for many people, evaluation of their predicted outcomes was straightforward for the practitioners with whom Colla worked. The models reference standard financial tools that are easy to understand when evaluating a business case, and likelihood distributions are more flexible and compelling than static assertions. A focus on NPV portrays a time-based and more nuanced picture of ROI than does a more standard use of ROI alone.

The results so far suggest that a good business case can be made for small Argentinean organizations to find value from process improvement even at lower maturity levels. A good business case can be made that medium-sized and larger organizations in Argentina can expect returns on their investments while achieving higher levels of maturity. Moreover, benefits are likely for organizations operating in volatile and uncertain environments. Still, there is a critical dependency on allowed investment horizon that is greater than what has been perceived as affordable by Argentinean small and medium enterprises. Support from government or cooperative efforts for hedging

instruments such as deployment frameworks to reduce cycle time most likely are necessary to help them receive payback on their investments in process improvement.

Evaluation of SPI Efforts in Small & Medium Organizations: An Update with New Results

Pedro E. Colla, Instituto Universitario Aeronáutico and Universidad Tecnológica Nacional-Facultad Regional Santa Fé; and Jorge Marcelo Montagna, INGAR-Instituto de Desarrollo y Diseño, Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional-Facultad Regional Santa Fé

In this presentation Pedro Colla provided an update on the work described in "Evaluation of SPI Efforts in Small & Medium Organizations" on page 28. After reviewing the work and introducing it for those who were not at the previous workshop, Colla described some new results and provided an update of the project's status.

This work is meant to help small- and medium-sized software development organizations throughout Argentina better understand the costs and benefits of investing in software process improvement. Rather than focus on a single organization at a time, the modeling is based on comparisons across organizations. The modeling effort was both deterministic and stochastic. The deterministic model is based on mathematical transfer functions meant to integrate the analytical work into a single framework (using VenSim[6]). Colla briefly described that model; however, he once again focused on the stochastic results of Monte Carlo analyses (using GoldSim[7]) in this presentation. He also briefly discussed some of the results based on survey data collected expressly for the modeling effort.

Most organizations in Argentina and South America that wish to participate in software development projects in international off-shore markets are small or medium in size. The model predictions are meant to help policy makers in those organizations as well as the Argentinean government better understand what drivers facilitate adoption of process improvement practices by such organizations in order to increase their competitiveness in both their domestic and off-shore markets. The model results suggest that a good business case can be made for small organizations to seriously consider CMMI-based process improvement initiatives, and medium organizations appear to have good reason to aim for high maturity status. However, the time necessary to achieve payoff for investment in process improvement often is greater than what typically has been perceived as affordable. Hence Colla and Montagna are examining how hedging instruments made available with government support and through collaborative initiatives may help such organizations share infrastructure costs. This is especially important for organizations that operate in volatile and uncertain markets where the benefits of process improvement need to be produced as early as possible.

Colla presented four new sets of model results during this presentation. Results from a field survey among all Argentine information technology organizations provided some useful background when interpreting the model results. Sufficient data for analysis were provided by 111 organizations (64% were focused on software development and maintenance). Colla also pointed out that

---

[6] Further information about VenSim is available at http://www.vensim.com/.

[7] Further information about GoldSim is available at http://www.goldsim.com/.

the modeling was better calibrated with the field data for the results reported in this second presentation, and the model predictions were consistent with the comparable survey results. The data points used for the calibration were from Argentina only; however they appeared to be consistent with similar markets.

As described in his first presentation, Colla used net present value (NPV) as his major y factor because it summarizes both cost and benefit cash flows and discounts them appropriately based on each organization's cost of capital in dollar terms and accounts for the time value of the money invested. That is especially important under risk-averse circumstances when current status can trump uncertain futures.[8]

As shown in Figure 7, the proportion of organizations that have begun CMMI-based improvement initiatives rose considerably along with organizational size as did the probability of realizing a positive net present value from investments in process improvement through achievement of CMMI maturity level 3. All of the organizations analyzed were included in the figure, regardless of their current maturity levels. The vertical bars (FREQ [%] ARG) correspond to the proportion of organizations adopting CMMI from the total number of organizations in each organization size bucket across the x axis. The lines (tp) correspond to the proportions of organizations in each size bucket that have achieved a positive net present value within each of three "investment horizon" time periods.
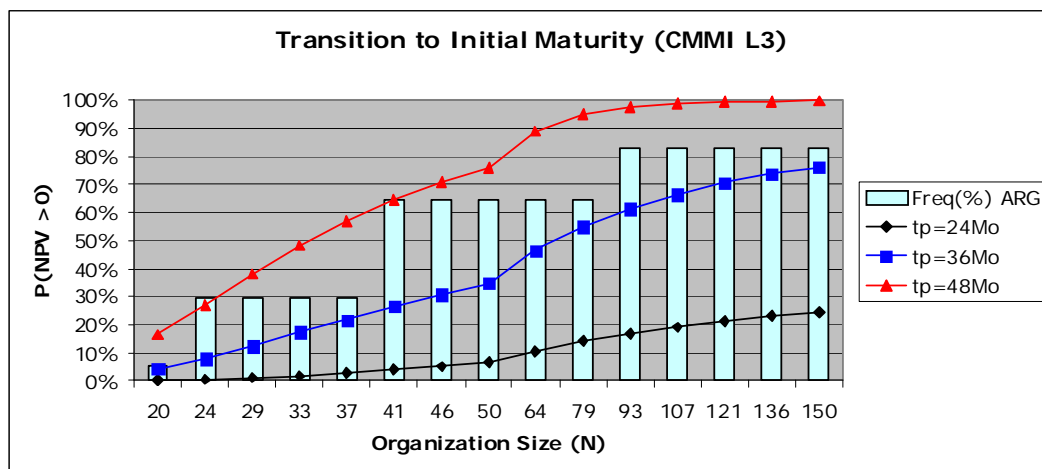
Figure 7:   NPV Through Maturity Level 3 as a Function of Organization Size and Time Horizon

Note that the probably of achieving a positive net present value within 24 months remains low regardless of organizational size. The rise is considerably higher after 36 months. It is not until a year later that the likelihood of achieving a positive net present value becomes substantially higher, especially for the relatively larger organizations, yet 85% of the surveyed organizations seek a time frame of 36 months or less. Moreover, the modeled likelihood of achieving a positive net

---

8   The commonly used ROI cost-benefit ratio is easy to compute and comparable across disparate projects. However, since it does not account for the time value of money (that the money may be worth less later versus when it was invested), there is no way to know the magnitude of the investment or return from the ratio alone, and there is no way to tell from the ratio how long it will take to achieve sufficient return to justify the investment.

present value is a fairly good predictor of whether or not an organization will adopt a CMMI-based improvement initiative.[9]

In fact, very few of the organizations surveyed had adopted CMMI at the time of the survey. For that matter, relatively few had sought ISO 9001 certification. ISO had been adopted by 41 of the 111 survey respondents (37%), but only 29 (26%) had adopted CMMI. As seen in Figure 8 however, organizations with 40 or more employees were much more likely to have CMMI-based improvement initiatives than those with fewer employees.[10] The latter were more likely to prefer ISO, perhaps because of a belief that CMMI may yield higher productivity benefits per person but requires a higher investment over a longer timeframe.
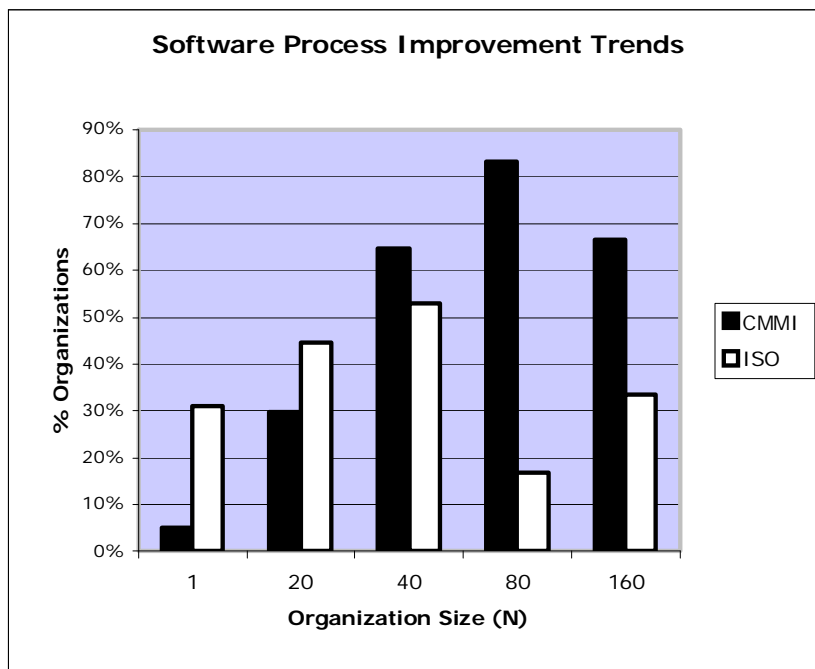
Figure 8:   Process Improvement Trends

Figure 9 depicts the probability of realizing a positive net present value from investments in process improvement through achievement of CMMI maturity level 5. Similar to Figure 7, the y axis again corresponds to the proportion of positive net present value predicted by the Monte Carlo modeling for each combination of organization size and investment horizon. The first three lines (tp) correspond to the proportions of organizations in each size category across the x axis that have achieved a positive net present value within each of the same three investment horizon time periods shown in Figure 7 (24, 36, and 48 months respectively). Unlike Figure 7, a fourth line in Figure 9 also depicts the frequency of organizations in each size category that have adopted CMMI-based process improvement initiatives aimed at achieving high maturity status (FREQ [%] CMMI).

---

[9] The quantitative result was not shown in the presentation.

[10] The Ns are small but the differences are statistically significant. There is no significant difference by size for ISO adoption.

Several differences between the two figures are worthy of note. First, hardly any of these organizations are likely to achieve CMMI maturity level 5 and also realize a positive net present value within a 24-month investment horizon, regardless of organization size. Not surprisingly, the proportion of organizations that are likely to surpass a break-even net present value within 36 months of their journeys to maturity level 5 status is noticeably lower than what they can reasonably expect on their way to meeting the goals of CMMI maturity level 3; however, a number of medium-sized organizations are likely to be successful within a three-year time period. Moreover, the Monte Carlo modeling predicts that most of the medium-sized organizations and even some of the smaller organizations can reasonably expect to realize a positive net present value while reaching maturity level 5 status within a four-year period. Notice also that all of the organizations that have adopted CMMI-based process improvement initiatives aimed at achieving high maturity status have 50 or more employees. Not surprisingly, adoption of high maturity initiatives also is correlated with the modeled likelihood of achieving a positive net present value.[11] Unfortunately, the small number of high maturity organizations in Argentina reduces the degree of freedom necessary for further analysis.
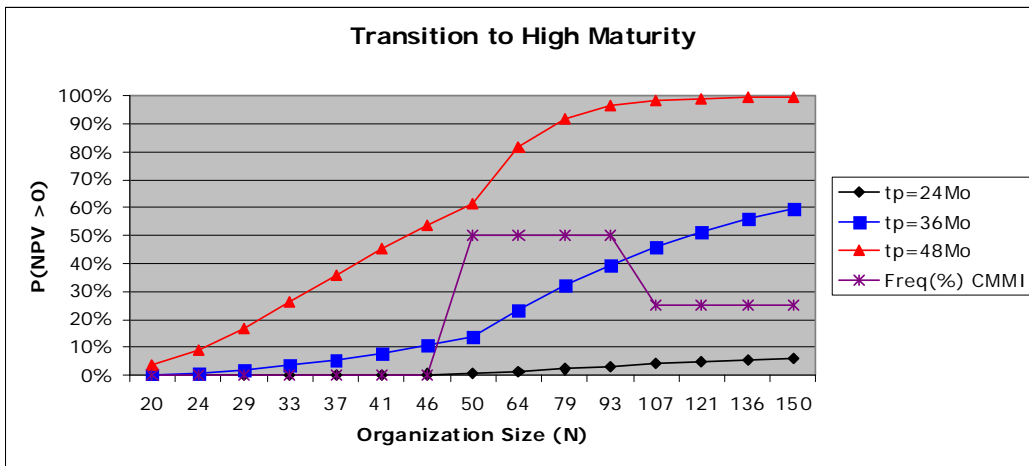
Figure 9:   NPV Through Maturity Level 5 as a Function of Organization Size and Time Horizon

Figure 10 shows the results of another Monte Carlo scenario that Colla modeled. As usual, the probability of achieving a positive net present value from investments in process improvement is depicted along the y axis. The probability of achieving each of several ROI cost-benefit ratios is shown across the x axis. The lines represent each of three categories of cost allocated per engineer (CPE) by the modeled organizations. The ROI ratios are constrained to be no less than 1:1; however, the ROI ratios of the underlying business and NPV dollar values from the SPI initiative are defined independently. While net present value is extremely important for organizations that are apprehensive about investing in process improvement, the amount of benefit they are likely to achieve also remains an important consideration.

---

[11] The quantitative result was not shown in the presentation.
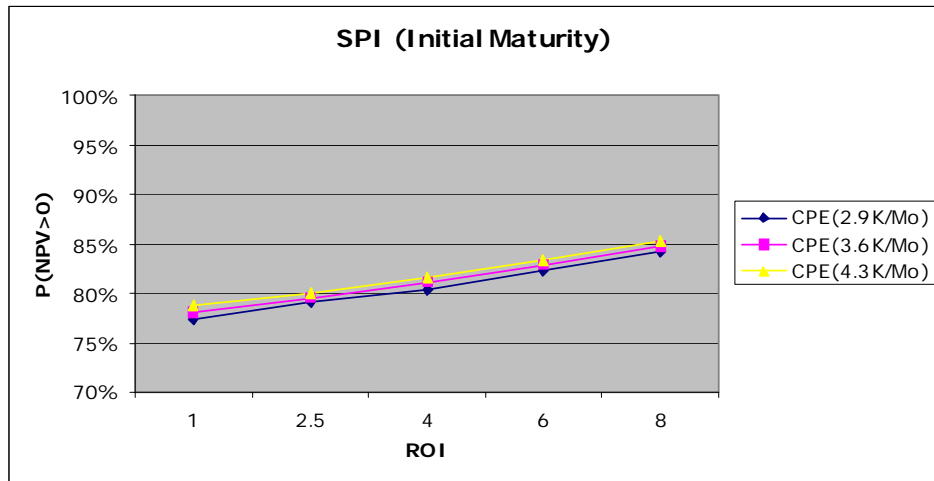
Figure 10: NPV as a Function of ROI and Expenditures per Engineer

Notice that the predicted likelihood of achieving a positive net present value does indeed rise with ROI. It also is slightly higher for those organizations that expend more per engineer. Colla pointed out that the latter is somewhat paradoxical since adoption of CMMI-based process improvement initiatives appears to be more common in emerging economies with lower costs per engineer. He conjectures that the expectation that a higher ROI will increase the likelihood of a positive net present value will encourage organizations with lower costs per engineer to consider adopting CMMI-based improvement strategies.

Colla also discussed another Monte Carlo scenario that examined the relationship between the likelihood of achieving a positive net present value and predicted return on investment in process improvement. Figure 11 shows that relationship as moderated by differences in business risk and CMMI maturity level. Business risk was estimated in the modeling based on the kinds of projects undertaken by the organizations as well as the environments in which they worked. The x and y axes are identical to those in Figure 10. The lines depict the model prediction for low, medium, and high business risk. The risk reduction factor (λ) in the figure legends is modeled as a function of certainty in process execution, where less variation in schedule and cost process outcomes is symbolized as lower risk.[12] The graph on the top of the figure (labeled "initial maturity") shows how the extent of business risk is predicted to moderate the relationship between NPV and ROI for maturity levels 2 and 3 organizations; the bottom graph (labeled "high maturity") shows the same relationship predicted for maturity levels 4 and 5 organizations.

---

[12] The risk reduction concept is based on work by Harrison, Settle, and Raffo [3]. The model parameter values for Low (0.6), medium (0.8) and high (1.0) risk were calculated based on data from Lawless, Flowe, and Thordahl [4]. See Colla [5] for more detail.
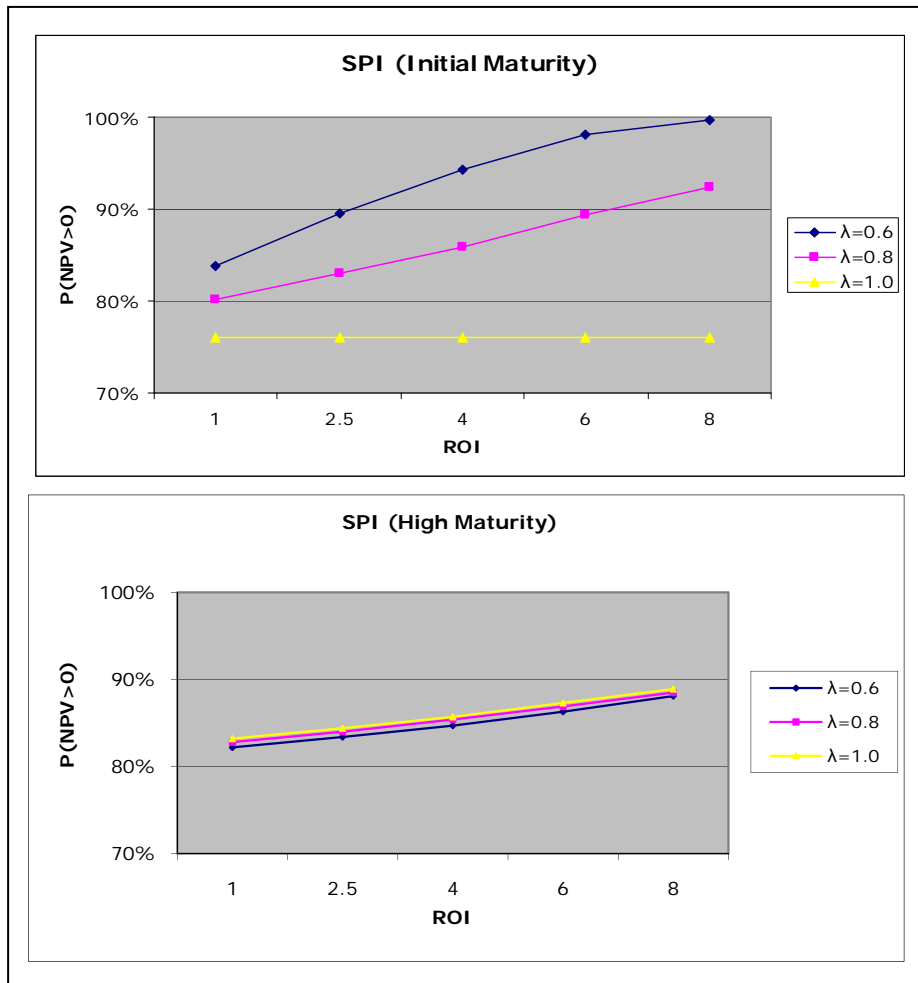
Figure 11: NPV as a Function of ROI, Business Risk, and CMMI Maturity Level

Colla's hypothesis was that lower risk should drive down opportunity costs, which in turn would yield a higher net present value. Notice however that the relationships differ considerably as a function of maturity level.

The predicted NPV value remains unchanged regardless of ROI for maturity level 2 and 3 organizations when risk is high, but it rises considerably along with ROI in the presence of medium and especially low business risk. The probability of achieving a positive NPV rose noticeably along with ROI for the level 4 and 5 organizations, yet the differences due to business risk were negligible.

Colla noted somewhat of a paradox in these results. Since lower business risk factors appeared to increase the likelihood of getting a good return from investment in process improvement at maturity levels 2 and 3, Colla conjectured that skeptical organizations might be willing to invest in reaching CMMI maturity level 3 but remain hesitant about high maturity initiatives. That might be so even though higher maturity organizations appeared better able to mitigate the risks they faced. High maturity organizations appeared to achieve their business value through other mechanisms

such as operational efficiency and productivity,[13] but that might be less compelling to organizations whose concerns are focused on risk.

In summary, Colla's modeling results suggested that CMMI-based process improvement efforts in Argentina were characterized by fairly fixed costs of deployment while the rate of return was very much dependent on organizational size. Therefore the relative "pain" was higher for smaller organizations. Government support and collaborative initiatives appear to be necessary to shorten the time it takes for such organizations to get the results (tp) that they need to justify the investment. That said, the modeling results also suggested that organizations with lower costs per engineer might be encouraged to adopt CMMI since they appear to have an opportunity to achieve higher returns on their investments in process improvement than do their counterparts in more developed markets. Similarly, the value added through risk reduction mechanisms seems to have a higher impact when Argentinean organizations achieve maturity levels 2 or 3, while gains are achieved by higher maturity levels through other mechanisms such as operational efficiency and productivity.

2.3   Other Simulation Approaches

Conceptual Planning, Execution, and Operation of Combat Fire Support Effectiveness: A Thinking Model with Practical Measurements

Kobi Vider-Picker, K.V.P. Consulting

In this presentation, Kobi Vider-Picker demonstrated the use of Bayesian belief networks (BBNs) and process flow simulation models for the definition of end-to-end life-cycle processes that require coordination among disparate stakeholder groups to meet product quality objectives and use resources efficiency. The presentation included evidence of all of the healthy ingredients for PPMs.

Vider-Picker described the use of process performance modeling as part of a larger quantitative management process for combat fire support. After describing the larger problem, he focused the discussion on the application of Bayesian networks and process flow simulation. Since the work on combat fire support was ongoing and could not all be discussed publicly, he concluded the presentation with an example of an industry product system that he could describe more fully. The modeling was strongly influenced by results from work by Carbonell-Foulquié and others [6].

The presentation showed how Bayesian networks can be implemented to provide decision support for the development and management of complex systems, notably including strong support for proactive what-if scenarios. The implementation of an intuitive graphical user interface can hide the complexities of the Bayesian network. In addition to quantitative data, such models can incorporate qualitative and expert judgmental data and handle missing data gracefully.

Vider-Picker was asked to evaluate a major fire support process for the Israeli Army. The work included other evaluative approaches including game theory, quality function deployment, and elements from the SCAMPI appraisal method.

---

[13] Quantitative results were not shown in the presentation.

Processes for effective planning, execution, and operation of combat fire support can be extremely complicated and challenging. Major operational challenges included the following:

- fire elicitation to achieve combat mission statements
- operating fire support with appropriate timing
- target planning
- planning and developing the fire support array
- evaluating enemy status and comparing it to the mission objectives
- coordinating everything end-to-end from target definition and identification to ammunition availability and logistic considerations
- developing a support plan, including acquisition planning, warfare research and development (R&D), adjusting ammunition to scenario, and scenario simulations and testing

Fire support units prepare the battlefield for the direct assault ground forces. Effective battlefield management is dependent on the commanding officer's capability to accurately evaluate the fire support unit's effectiveness and the efficiency of its resource usage. To be deemed useful by the customer, the military mission objective statement had to include quantitative objectives that were stated in a clear way, ensuring that the executing force unit and its command would be able to quantify the achievements of their objectives.

Working with several stakeholder groups to translate these problems into a controlled process, Vider-Picker and his team identified the following operational needs:

- mapping and classification of targets to operational priorities
- adjusting single targets to operational achievement
- adjusting target life-cycle time to attack timing
- adjusting ammunition to target profile
- adjusting ammunition elicitation to target profile and mission success objectives
- adjusting ammunition to combat platforms
- determining the platforms' accessibility to target and target life-cycle time

Major methodological challenges for an effective solution included the following:

- information analysis
- target structure analysis
- target position in the destination environment
- target value chain
- operational system value chain
- weapons elicitation to target type and classification

Major operational challenges for the solution included the following:

- definition and structuring of mission objectives quantitatively
- definition of a "good enough" level
- differentiating mission objectives and success factors for different battle phases
- resource usage and elicitation of adjustments to plans and objectives

All of this need for planning, execution, and control was further complicated by the need to coordinate among several multi-functional teams and groups that were in different physical locations and organizational units. Hence the overall collection of parameters required a process that considered multi-dimension relationships and the impacts among core entities and their building blocks and on the decisions that needed to be made.

Based on these and other considerations, the team developed a Bayesian belief network[14] to illustrate the relationships and dependencies among main entities of platforms, ammunition, targets, timing, intergroup coordination, and mission objectives. A simplified illustration of the overall conceptual mode is shown in Figure 12. As shown in

Figure 13, each core element (e.g., ammunition, target, timing, and required achievement) exists in its own three-dimensional conceptual space, and some share overlapping space with two or more other core elements across the three dimensions.
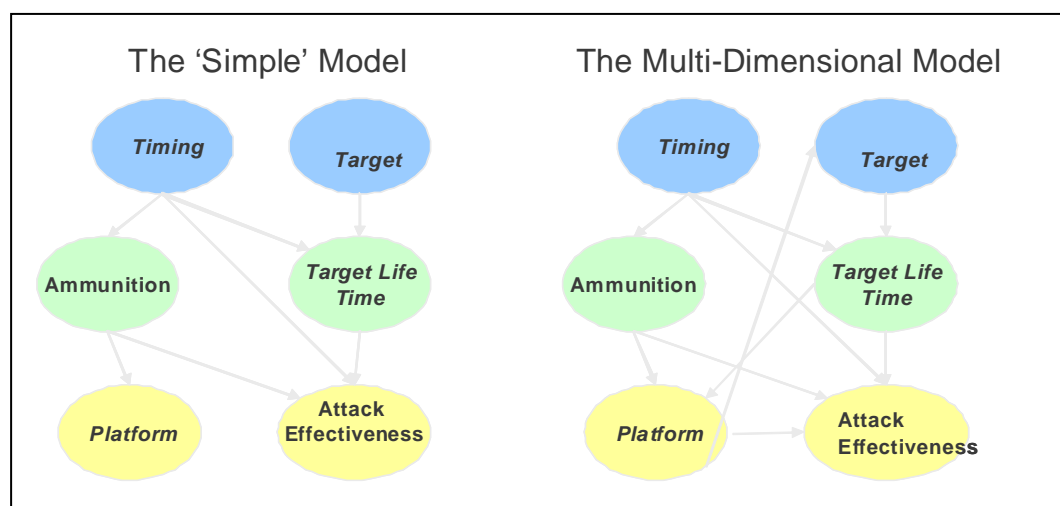
Figure 12: Bayesian Conceptual Way of Thinking

---

[14] Bayesian networks are directed acyclic graphs whose nodes represent variables and whose arcs encode conditional independencies between the variables. The nodes can represent any kind of variable, be it a measured parameter, a latent variable or hypothesis. The nodes are not restricted to representing random variables, which is another "Bayesian" aspect of a Bayesian network.

The team used Hugin, which is a Dutch Bayesian Belief Network (BBN) tool. Others, including Netica and Agena Risk are less expensive; however Hugin was available through a customer with whom Mr. Vider-Picker works.
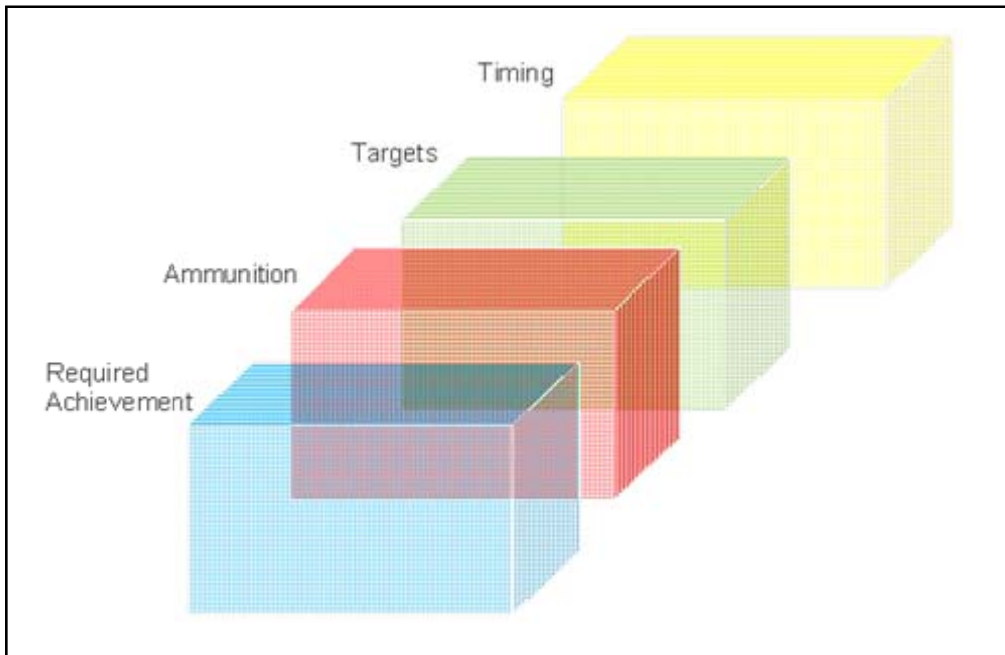
Figure 13: Three-Dimensional View

After summarizing the essentials of the fire support example, Vider-Picker presented further details based on similar work done in a civilian context by describing the development of the model and the way it was used by the organization for which it was developed. The challenges faced there were very similar to those in the military example. In both instances, there was a need to model an end-to-end life cycle rather than a limited, single phase. The focus in the presentation was on a new four-stage product-development process. The factors being modeled included independent evaluations of the managers' capabilities to evaluate the effectiveness in meeting product quality objectives and efficiency of resource usage in the organizational units for which they were responsible. The model was based on the company's approved life-cycle phases.

The business needs faced by management included the following:

- mapping and classification of requirements to development and operational priorities
- adjusting single requirements to achievement of developmental objectives
- adjusting component life-cycle time to maintenance timing
- adjusting development and maintenance methods to required component profiles
- adjusting development and maintenance methods to system and subsystem profiles and success objectives
- adjusting development and maintenance methods to the system's operational environment

The simplified development process was used to predict the potential return at four selected stage gates: new product concept, new product design, production startup, and subsequent market return. As seen in Figure 14, the modeling team identified and selected thirteen relevant criteria and grouped them into five main factors. The criteria, factors, and process stages to which they map are shown as nodes in the Bayesian network. The arcs in the figure from specific criteria to their

parent factors indicate the hypothesized causal relationships between them (e.g., sales growth and market share influence market opportunity). Similarly, the arcs between the five factors and the four stage gates indicate that each factor influences the return at each stage gate.[15]
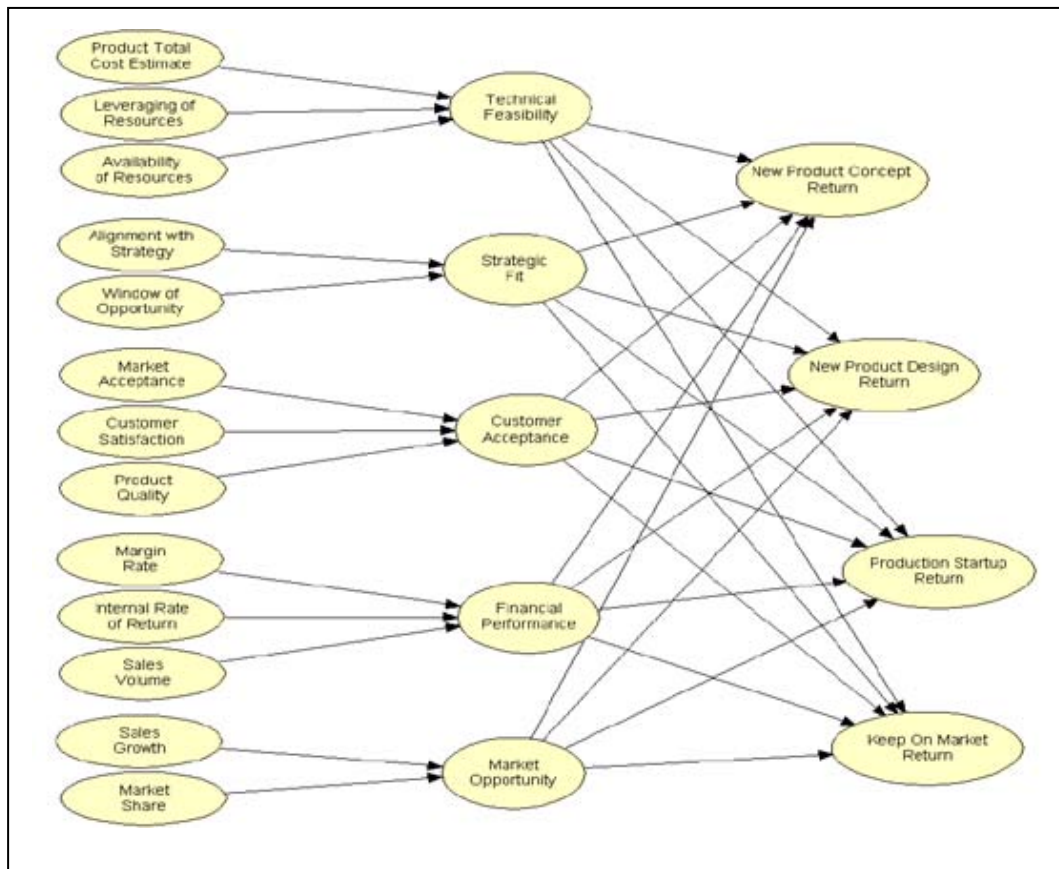
Figure 14: Bayesian Representation of New Product Development

Once the relationships were established for the network as derived from the baseline data, the next step was to associate probabilities with the causal relationships. Doing so required defining appropriate states for each of the nodes. Since there were so many, the team decided to use three states for each of the criteria with numerical intervals ranging from 1 to 3. These states were interpreted from worst to best for each of the criteria and then normalized so that the factor values would always range between 0 and 1. Doing so simplified interpretation for the users of the model outputs as well as the development of the expressions for determining the probabilities associated with the new product development (NPD) return nodes.

The states for the NPD return nodes were determined by the possible states for the factors and the weightings of the relationships. It turned out that a granularity of only three states to represent the NPD return did not provide sufficient resolution to aid the user's understanding of the results. Thus the team decided to implement four states for those nodes, which were referred to as low, medium-low, medium-high, and high returns.

---

[15] One of the assumptions for the causal relationships is that the criteria that influence a factor do not change between the four new product development stages.

An important benefit of the Bayesian network is that the value for any node is not limited to the input from earlier in the causal chain, in this instance from the criteria nodes. Evidence can be entered at any of the nodes and will propagate through the network. Thus all manner of what-if scenarios can be entertained to further structure and refine the elements of the decision model to facilitate process improvement. Several model outcomes were described in the presentation. For example, if a higher level of customer satisfaction could be achieved, the model predicted that the probability of receiving a high return at the design stage would almost double (from 17% to 31%). Similarly, the probability for high return changed from 13% to 24% at the production startup stage, and the 12% probability of achieving only a medium-low return for the stage after market release would disappear entirely. Similar results were shown for better ROI if resource availability increased.

Game Theory and Bayesian Belief Network to Support Quality Function Deployment (QFD) for Requirements Development

Kobi Vider-Picker, K.V.P. Consulting

This presentation showed how game theory and Bayesian probabilistic methods were used to predict program success based primarily on stakeholder and engineering factors. All of the healthy ingredients were shown in the presentation, with the exception of controllable factors, connecting upstream with downstream activities, and enabling projects to make mid-course corrections.

Requirements development and management remain major challenges for the development and sustainment of complex software intensive systems and innovative new products. Not only do such efforts face difficult engineering problems, those problems often are compounded by multiple stakeholder groups with very different agendas. In this presentation, Kobi Vider-Picker described several advanced decision techniques that may be effective under such circumstances. His approach was developed during a year and a half working with the Israeli Air Force on a joint development life-cycle process where it became apparent that a clear way was needed to balance different stakeholder agendas, operational needs, operational scenarios, and requirements. The project involved internal, external, and contractor stakeholders.

Vider-Picker was asked to develop a method that would support the initiation of complicated projects where a large number of overlapping stakeholders influence the scope of a system, product, or program and its deliverables. Communications flow can be quite difficult under such circumstances, especially with large, complex, or global teams. Competing expectations often get lost. A common structure or logic for resource allocation remains unclear or lacking altogether. The same is so for teamwork coordination, and the team members and teams change over time. All of this results in difficult challenges for establishing efficient and effective processes along with excessive redesign, repetitive problem solving, and fire fighting.

Business goals were set for the project to address these difficulties. The goals included the following:

- simplify each development initiative by clarifying its scope, intended uses, and users
- identify, map, and assign appropriate priorities and commitments for the different stakeholders
- identify and predict the new product initiative or invention's impact on the program and other stakeholders

- identify and predict the impact of team coordination and alignment efforts on the program and the different teams or team members
- identify and predict the impact of changes in process efficiency and effectiveness on the program and teams
- identify and predict conflicts in development time with stakeholders expectations
- identify and predict the effectiveness of redesign on the program and teams
- identify and predict the impact of team changes on the program and teams
- identify an effective way to choose between and improve methods for proactive problem solving and fire fighting based on quantitative and predictive impact analysis

Vider-Picker described how this combination of game theory, Bayesian networks, and quality function deployment (QFD) used in a highly complex system can be translated into a simple yet multi-dimensional quantitative process for requirements development. It can be used intuitively by management to reduce the time required for achieving mission goals and deliverables while balancing stakeholder needs.

Vider-Picker's team did postmortem, retrospective analyses of five programs to establish a baseline for evaluating methods to determine the most appropriate one for their customer's purposes. These included classical game theory, QFD, Bayesian networks, dynamic Bayesian games, and related voice-of-the-customer impact analyses. The presentation included summaries of the pertinent process elements that they were able to identify and the parameters they used for performance measurement. The presentation also included some detailed discussion of the tools that could be used and made available to the workshop participants.

The effort started by mapping the stakeholder preferences and assigning different weighted gains and losses for the stakeholders associated with engineering options at different life-cycle stages. The result was a quantifiable, measurable matrix of x factor[16] parameters along with guidelines for its use. The matrix was used in the QFD sessions, as the baseline for simulations and for inputs to the other decision-making and influence analyses.

Much of the work was done with spreadsheets. The initial stakeholder mapping and analysis was mapped in one Excel spreadsheet and the aspects of the engineering options were mapped in a second spreadsheet in the same workbook. The team also developed a detailed spreadsheet for their QFD work and another for the classical game theory analyses. They developed two Bayesian networks. One was for stakeholders' influence analysis and included the stakeholder groups, life-cycle phases, and expected outputs from each phase. The other included aspects of the engineering interfaces and their influence on phase delivery success, overall impact on product readiness, and operational compliance. The Bayesian network files were built in Hugin. At the time of the presentation, Vider-Picker was considering converting it to be used with additional tools.

Due to the unique nature of data elements and related factors, the data elements and factors were managed manually based on stakeholders organized by project or program. At the time of the presentation, the team had initiated an historical database and was in the process of developing a

---

generic data model. No sampling was done since they needed to run the full method from the start of each project or program. The team faced a number of threats to data quality and integrity, including stakeholder subjectivity, unclear role assignments, and changes of people in the same position over the course of the analyses. At the time of the presentation, the team was running a postmortem analysis on their past projects to clean the existing data and better understand the extent of measurement error.

The presentation also contained descriptions of the various modeling techniques that the team used. They considered game theory because they recognized that the tradeoffs among their stakeholders were often competitive. That was the intent when this approach originally was developed by von Neumann and Morgenstern in 1947. However, classical game theory cannot predict optimal solutions under circumstances involving competitive judgments made under uncertainty. Solutions can remain undetermined or predict self-defeating behavior that is contrary to empirical results. While classical game theory is useful to understand social interactions, it became clear to the team early on that it needed to be modified for their purposes.

The team chose to use Bayesian belief networks to better understand decision-making processes for establishing project and program scope. It often is difficult to explain how decisions relate to and affect architecture considerations. The same is so with respect to quantifying the impact of changes in requirements and design. Moreover, the customer needed to consider multiple factors including non-functional requirements for quality attributes and environmental factors as well as functional requirements. The customer also wanted to consider different perspectives and viewpoints that were not always heard. Moreover they wanted to more directly and indirectly influence system design structure to satisfy system goals and subgoals, and they wanted to encourage early decision making when changes were needed.

The team chose Bayesian games since they can deal with incomplete information and sequential bargaining. Another important consideration was that cooperation could be sustained as a Nash equilibrium when there was no certain end to a game or when there were multiple equilibria. Doing so required the ability to monitor actions of rivals, the ability (and reputation of) punishing defectors, low interest in particular decisions, and high probability of future interaction—all of which characterized the problems faced by the team's customer. In dynamic Bayesian games with asymmetric information, the players also can learn about other players through actions that are chosen before they themselves have to make decisions. This too characterizes the circumstances with which they had to deal.

Vider-Picker briefly described the history of quality function deployment and its house of quality approach in conjunction with Kano analysis and its focus on customer dissatisfiers, satisfiers, and delighters. These methods, which are not yet widely used in CMMI-based process improvement, also hold great promise for improving requirements development and change management in complex software intensive systems. He also briefly described the team's use of decision analysis, another approach that may provide effective methods for organizing a complex problem into a structure that can be analyzed by identifying possible courses of action and possible outcomes along with their likelihood and eventual consequences. Decision analysis also provides methods to analyze tradeoffs of benefits against costs, and it may help people make effective decisions more consistently.

In the end, the team's approach was well received by the customer. Senior staff commitments were maintained, and the various stakeholder groups accepted the importance of balancing their influence with others to achieve better results. The stakeholders also accepted the weights assigned to them in the now explicit decision-making process. They liked having a clear view of all aspects of the work, and they appreciated how an analytic modeling approach reduced the complexity of decision making. The use of the historical database from past projects also reduced resistance to the new approach. Other side benefits included other departments expressing a desire to be included in future work, requests for the development of a generic model for use elsewhere, and a request to adjust the overall model for strategic and multi-year programs.

Product Modeling to Insure Reliability of High Maturity Indicators

Jim Perry, BAE Armament Systems

In this presentation, Jim Perry used algebraic, probabilistic, and Eigenvalue structure matrices to enable the prediction of impacts on software quality attributes as a result of changes to detailed subprocesses tightly coupled to software and system components. All of the healthy ingredients for PPMs were evident except for the modeling of uncertainty or variation of the factors and outcomes, which could be handled with extensions to what was shared.

The presentation was based on five years of process improvement work on a very large development program involving software, systems, mechanical, electrical, controls, and specialty engineering. The program used the Software CMM, CMMI Version 1.1 for Software, and CMMI Version 1.2 for Software and Systems Engineering. The focus of their effort was to improve program performance and delivered products consistent with organizations operating at CMMI maturity level 5 for software and level 3 for the other engineering disciplines.

In the presentation, Perry described an uncommon approach to process performance modeling where the model elements were tightly coupled to detailed aspects of the structure of the product being developed. This design model was being piloted in a major development program at the time of the presentation to address limitations of their current models. Their process performance models had been based on organizational processes—primarily review subprocesses—and were used mostly by process engineers for process improvement. Yet at the project level, their goals addressed the products they built, and they wanted a model more closely connected to that. While they recognized the importance of process improvement, they wanted a model that would be more useful for development engineers making development decisions.

Perry began by showing a magnificent digitally enhanced photographic print of reflections on Walden Pond. The image was created using a controlled chemical process for dye transfer where the quality of the image is determined by an artist's specifications and targeted sale price. According to Wikipedia, dye transfer "allows the practitioner the highest degree of photographic control compared to any other photochemical color print process." A similar approach for Perry's design model is based on earlier work in reliability engineering by Rosene and associates published in 1981 [7].

The business and project goals that the design model aims to facilitate are stated explicitly in a hierarchy. The group is responsible for improving cost and schedule estimation, quality, and productivity. The program is responsible for technical performance measures and CPI, and the project is responsible for managing requirements and design changes. Model alignment and use

were similarly structured for any specific goal: the higher level organizational group was ultimately responsible for development decisions, the program was responsible for timing and reliability, and the project was responsible for decisions related to modifiability and interfaces based on the product structure. The design model was used to support design, integration, and verification and validation decisions for all of these stakeholders. Goals were set for intermediate products to minimize the risk of achieving higher level goals later in the product life cycle. The model was meant to help identify the conditions under which positive or adverse impacts were likely to occur so development activities could be better aligned to achieve the goals at all levels.

The primary outcomes predicted by the model were indicators of the impact of product and process changes on the interim and deliverable work products aimed at controlling product complexity, modifiability, and reliability. Additional outcomes, expressed as ordinal data, were used to guide decisions about scheduling and prioritizing verification and validation activities. Software designers and testers were largely responsible for the product development decisions, while process engineers took the lead for process improvement decisions.

Process was aligned tightly with product. The model captured product attributes such as complexity and size as they apply to design, implementation, integration, and test. That context was used to support and capture development decisions, and process improvement was measured by product improvement.

The design model was algebraic and probabilistic with a structure based on linear and matrix algebra. Several qualitative factors were used in constructing the model. The product structure was determined by architectural rules, all of which were nominal and controllable. These included hierarchy levels, dynamics activation, data isolation scoping, and communication messages. Algebraic model factors were based largely on historical data from the project's incremental development releases and builds, which were augmented with a limited amount of published industry data. Continuous, ratio data were used for the model's i, j matrix values. Eigenvalue multipliers and determinants were expressed with continuous values, but they were used nominally for interpreting the results. Measures focused on amount of communication, activation connectivity, and amount of data sharing. These were represented in control matrices for configuration items and data interface matrices for the operating environment. Both component items and data items were included in structure matrices. Design measures included communication and activation connectivity as well as amount of data sharing among configuration items in their respective operating environments.

As described more formally in Figure 15, probabilities that a change in each component i in a structure matrix will cause a change in each component j were estimated based on data and expert judgment for each matrix cell (e.g., probabilities were based on relationships among actors and associations in use cases). Components and data were grouped by level based on the product structure, and estimates were made for each level in the product structure. Expected total changes to the components and data were calculated based on the number of changes to all of the ith components that were estimated to affect each jth component. Hence the model supported what-if analyses based on anticipated and suggested changes to the matrix cell estimates.

- Mij is the probability that a change in component i causes a change in j

M		C	D
	C	p_{cc}	p_{cd}
	D	p_{dc}	p_{dd}

- Estimates of the probabilities
  - Components and data are grouped by level per product structure
  - Estimates made for each level in the product structure
- Expected total changes to components and data, $T = A(1-M)^{-1}$,
  where Aj is the number of changes to component or data j
- $(1-M)^{-1}$ exists if eigenvalues of M are > 0 and < 1

P is an n by n matrix, where i, j represents the probability that a change in component i will cause a change in component j

A is a 1 by n matrix where element j is the number of initial changes in component j

Then $A(I+P+P2+….+Pn+…) = A(I-P)-1$ is a 1 by n matrix where element j represents the predicted number of changes to component j

Figure 15: Interpretation of Structure Matrices

The structure matrices were used during development. As the structure evolves the matrix evolves. The estimated and predicted values were used in making decisions with respect to process performance, development, integration, verification, and validation. The final structure matrices provided reliability estimates for the final products that could be attributed to the tightly coupled process improvements used in building them.

The pilot project described in the presentation was begun to determine stakeholder value, obtain buy-in, and determine feasibility. Obstacles related to building the model included ongoing structural re-engineering, reliance on manual collection for the pilot, and limited resources available from the development engineers. The work also was limited by lack of consensus on measures for design, reliability, and maintainability. The model's scope covered both software and controls engineering, and the product was still under development with no delivered product as yet when the presentation took place. All of these obstacles also affected tradeoffs between full implementation of the matrices vs. over simplification.

In the end the model was well accepted by the engineers because it related directly to questions that they faced and the decisions they needed to make. It was not seen as needless extra work. Costs of model development were relatively low for a product development project of this magnitude. The measurement lead and one software process lead together spent about 300 hours

over a two-and-a-half month period, while the development engineers together spent about 20 hours per month. The costs of model maintenance were low: about eight hours of manual effort per month in the current phase at the time of the presentation and two hours per month for maintaining the automation. Qualitative benefits that were being recognized during the pilot included use of the model for independent verification of the then-existing design; support for improving reviews, integration, and test; new organization assets for development such as criteria for the development, test, and review processes; and better capture of development decisions to provide better context for business goals and related measures. Quantitative benefits had yet to be determined at the time of the presentation; however, new measures that were being defined included the number of reviews and test procedures changed or added, the number of development decisions supported, and more detailed counts of integration and test effort.

Software Maturity Modeling

Lynn Penn, Lockheed Martin Information Systems and Global Services

This presentation authored by Lynn Penn, Lockheed Martin Information Systems and Global Services, described the use of software reliability growth models to predict when a software product was mature enough to ship to the customer (i.e., to predict ship readiness using attributes and results of software testing). The presentation included evidence of most of the healthy ingredients, with the exception of controllable factors, what-if analysis, and connecting upstream activities with downstream activities. However, plans were in place to extend the modeling to include these ingredients.

Four of the criteria used in industry to make this decision are listed below.

1. The rate of code changes per defined time interval is less than a specified value.
2. All tests have been run successfully.
3. There have been no defects of the highest severity for a specified number of time periods.
4. Software reliability analysis indicates that remaining latent defects are below a specified level, and there are no outstanding defects of the highest severity.

Lockheed Martin chose an approach that combined the first and fourth criteria above. Maturity was based on the rate of change of defects detected during testing. When the rate of change was below a target value, the product was ready to ship—that is, the reliability of the software was acceptable. The software reliability analysis was the foundation for the decision.

Software reliability is the probability of failure-free software operation for a specified period of time in a specified environment. A software reliability growth model (SRGM) is the mathematical relationship that exists between the time span of testing (or using) a program and the cumulative number of defects discovered. A software reliability growth model may take one of two perspectives:

*   time between failures (which is closer in spirit to the definition of software reliability above)
*   interval counts (which uses the number of defects per interval, usually a fixed period of time)

The Lockheed Martin model is based on either the Yamada S-shaped curve or Weibull software reliability model. The assumptions listed below must be true.

- The software will be operated in a manner similar to that for which reliability predictions were made.
- Every defect has the same chance of being encountered within a severity class as any other fault in that class.
- The initial fault content of the software system is a random variable.
- A defect is corrected instantaneously without introducing new defects into the software.
- The software is operated in a similar manner as the anticipated operational usage.

Lockheed Martin uses two tools to conduct the reliability analysis:

1. Statistical Modeling and Estimation of Reliability Functions for Systems, Software, and Hardware (SMERFS)
2. Software Error Estimation Reporter (STEER)

SMERFS uses defects and test interval length as inputs. Consequently, the interval length can be interpreted as the number of testing hours in a fixed interval (e.g., a week) with the number of test hours varying from week to week. If only defect counts in each interval are available, the interval length is treated as 1. However, the best fit resulted from grouping hours and defects from consecutive weeks into intervals of approximately the same size based on a fixed number of hours (e.g., the number of defects in a 6000 testing-hour segment). An additional heuristic used by Lockheed Martin to perform a sanity test on a particular model prediction was to find out what percentage of total defects would be identified in the last three intervals. A value of 1-2 % was acceptable, while values above 2% were suspect and troubling.

Although not explicitly appearing in the currently used mathematical model, the following factors were treated as controllable by Lockheed Martin in attempting to influence the volume of latent defects predicted by the reliability growth models:

- increasing staff, average capability (learning curve penalty)
- increasing or changing staff, superstar capability
- changing test management and leads
- test environment changes that improve efficiency and/or effectiveness (e.g., automation, more systems for testing)
- changes in functionality delivered in working condition (e.g., less expected defects)
- changes in testing strategy (e.g., order of testing)

As such, future work may focus on explicitly linking these factors mathematically to the prediction of latent defects in concert with the software reliability growth models.

The current Lockheed Martin model is run at least monthly during formal testing using actual effort and defect counts as inputs, producing a prediction of cumulative defects and time period profile with prediction intervals through the use of non-linear regression. The projections are compared to the original plan so that present feasibility can be evaluated and corrective actions undertaken if needed.

Initial observations of the Lockheed Martin Software Maturity Model are listed below.

- The primary stakeholders, customer, and senior management were pleased with the approach and had more confidence in the status provided. (It is difficult to quantify "status confidence" except to say it contributes to "green" customer satisfaction ratings as well as award fee levels.)

- The mechanics of the tools were easy to learn.

- The tools were free.

- Input factor data was readily available.

- There was sufficient management and customer support for the modeling and predictions.

- Assuming defect counts, labor, and tools were readily available, the development of the model would require, at most, 40 hours to become familiar with the tool set and, at most, 20 hours to experiment with the models versus the actual data.

- Not more than five hours per month were required to maintain and use the models.

- The modeling produced much discussion among members of the program management team that resulted in an educational benefit and the beginnings of formulating alternate strategies should they be needed.

- There are many possible software reliability growth models. There are subtleties that should be understood to apply a given model properly.

- The Lockheed Martin modeling tools did not support all of the desired models (e.g., Gompertz).

- The modeling tools were fine for predicting the downstream implications of labor consumption and defects found to date, but had no ability to directly provide the impact of other controllable factors.

- Reliability growth did not immediately appear in the test data. Consequently, the first interval for modeling purposes may not be any of the first several intervals. The Laplace[17] test may be helpful and is being evaluated.

- SMERFS and STEER do not currently have very good documentation or built-in help (although early versions of SMERFS do have documents available for a modest fee).

In summary, the Lockheed Martin experience demonstrated practical business value from the use of software reliability growth models to predict latent defects and test sufficiency from a software maturity standpoint. The work presented in this case builds on a solid body of research, experience, and tool support[18] [9] [10].

Defect Discovery

Lynn Penn and Pete McLoone, Lockheed Martin

---

[17] The Laplace test is a method for determining whether there is a trend over discrete events in a process. The test approximates a standardized normal random variable (e.g., z-score). See [8] for a more complete description.

[18] See http://www.openchannelfoundation.org/projects/CASRE_3.0/ and http://www.slingcode.com/smerfs.

In this presentation, the authors described the use of Rayleigh curve fitting to predict defect discovery (depicted as defect densities by phase) across the life cycle and to predict latent or escaping defects. Although the role of overt controllable factors was not immediately evident in the model described at the workshop, the authors described ways in which the analysis could include controllable factors in the future. While data were available for historical defect discovery rates by phase of development, a way to determine targets for defect discovery by phase was needed for new programs whose characteristics are different from those used to establish the baseline performance. Lockheed Martin's goal was to develop a model that could help predict whether a program could achieve the targets set for later phases using the results to date.

Research indicates that defects across phases tend to have a Rayleigh curve pattern [11]. Lockheed Martin verified the same phenomenon even for incremental development approaches, and therefore decided to model historical data against a Rayleigh curve form using non-linear regression. Two parameters were chosen: the life-cycle defect density (LDD) total across phases and the location of the peak of the curve (PL). Modeling determined confidence intervals for both parameters. The model can be used to evaluate the impact of the program's characteristics on the LDD and the PL in terms of most likely, highest, and lowest values (50th, 95th, and 5th percentiles).

Inputs to the Phased-Based Defect Detection Planning model included

- LDD estimated range in terms of 5th, 50th, and 95th percentile
- PL estimated range in terms of 5th, 50th, and 95th percentile

Outputs included

- planned defect densities by phase with a performance range based on the uncertainty interval for LDD and PL
- estimated latent defect density

Inputs to the model during program execution included the actual defect densities by phase.

Outputs during program execution included

- fitted values for phases to date
- predicted values for future phases
- estimated LDD, PL, and latent defect density

Figure 16 depicts a Rayleigh curve phase-based example using STEER II.[19]

---

[19] STEER II is a proprietary Lockheed Martin tool. Chapter 7 in [12] has some discussion of the deterministic STEER I.
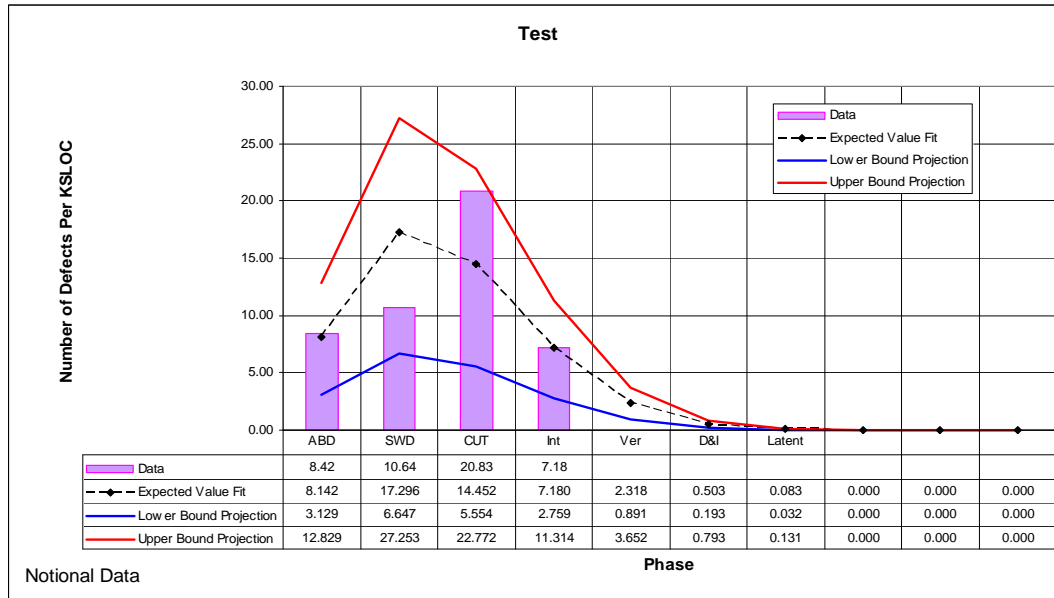
Figure 16: Rayleigh Curve Using STEER II

The presentation suggested that those who want to be successful using defect data by phase for managing a program need to

- be believers in the Rayleigh curve phenomenon for defect detection
- have an idea of what the life-cycle density should be as well as the peak location
- input their numbers into a modeling tool that can do non-linear regression to a Rayleigh curve and evaluate "what-ifs" until a plan is made
- fit their actual data to the same kind of curve and compare what is actually happening relative to the plan
- estimate peak location and life-cycle defect density accurately because they play a vital role in this type of modeling

Factors that are hypothesized to affect peak location include

- inspection preparation rate
- inspection pace rate
- staff skill levels
- proportion of the product representation inspected
- tool usage

Factors that are hypothesized to affect life-cycle defect density include

- product complexity
- staff skill
- defect detection methods
- requirements stability
- teaming arrangements
- multisite development
- tool usage

An additional regression-based model that would serve to quantitatively relate peak location or life-cycle defect density to measures of the above factors would be desirable and compelling. It could be used as part of an interconnected set of models that directly tie the controllable factors to the peak location and life-cycle defect density.

In conclusion, Rayleigh curve-based defect detection modeling using non-linear regression was an effective tool for planning and monitoring defect detection patterns to identify patterns requiring further investigation. As shown in Figure 17, the defect discovery phase-based model used the predictions from the STEER tool to establish upfront planning expectations as well as to provide updated predictions during program execution.
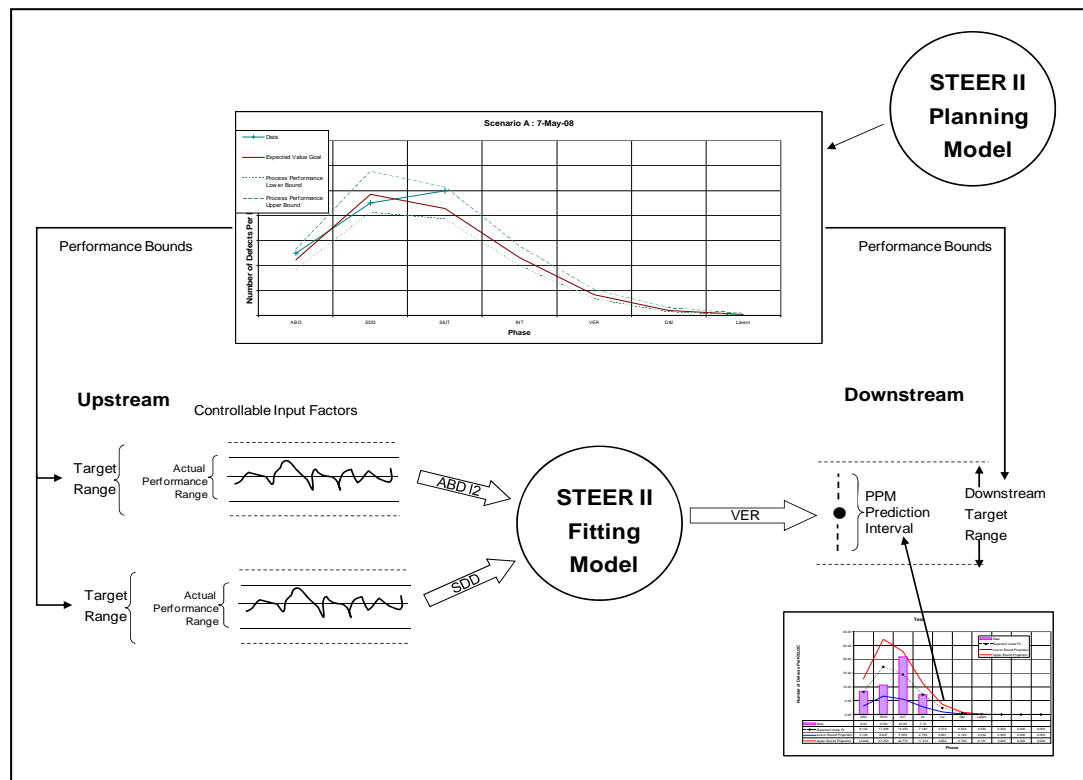
Figure 17: Defect Discovery Phase-Based Model Using STEER Predictions

For those considering this model, the SSCI SWEEP model is similar in intent, although deterministic in nature, and worth investigating. To perform the modeling described in this presentation, a tool is needed that performs non-linear regression using an iterative technique, such as the Levenberg-Marquardt method. Excel can do most of what is needed and can also be the foundation for a custom tool for this analysis.

Process Performance Modeling With Parametric Cost Models
Raymond L. Kile, PMP; John Gaffney, PE; Joan Weszka, Lockheed Martin

In this presentation Ray Kile discussed the use of parametric cost models (PCMs) modified to operate within the context of process performance modeling to predict cost, schedule, and quality. The modifications of the models and their usage included evidence of all the healthy ingredients except for the connection of upstream and downstream activities. The specific PCM discussed

was the COSYSMOR model and accompanying tool, developed for Lockheed Martin by John Gaffney of, which contains enhancements to the COSYSMO[20] model previously developed by Ricardo Valerdi at the University of Southern California.

Lockheed Martin uses PCMs for more than just forecasting project effort, cost, and schedule. They are also used to

- establish quantitative project objectives
- compose a defined process
- manage project executions

With that perspective, it was noted that models were needed at all levels of the development organization to support management decision making. Additionally, different disciplines would have different needs for models and different forecasting time spans.

Lockheed Martin pursues healthy characteristics of process performance models, meaning that the models

- relate the behavior or circumstance of a process or subprocess to an outcome
- predict future outcomes based on possible or actual changes to factors
- support what-if analysis
- use controllable factors from one or more subprocesses to conduct the prediction
- use statistical, probabilistic, or simulation models rather than deterministic models
- model uncertainty in the factors and predict the uncertainty in the outcome factor

The Lockheed Martin Parametric Cost Model (PCM) is defined as a mathematical representation that relates a cost/effort value to the values of various parameters. In an effort to correctly distinguish the traditional use of PCMs versus their acceptable use as process performance models, Lockheed Martin generated the following table to summarize the key distinctions.

Table 5:   Parametric Cost Models vs. Process Performance Models

PCMs	PPMs
Typically used "out of the box"	Typically developed locally for the domain of interest
No local calibration	Calibrated for the project
Not based on your processes	Based on your processes
Used only at beginning of program or at major milestones	Used frequently during the project and at major milestones
Compare actual to predicted performance at end of phase milestones misses opportunity to influence meeting objectives within that phase	Use "controllable x-factors" to manage process execution to achieve process and product goals and objectives

Through experience, Lockheed Martin discovered that a number of modifications to their traditional use of PCMs were required to embody the concepts of healthy process performance models:

---

[20]   See http://en.wikipedia.org/wiki/COSYSMO for more information.

- They should be tied to specific processes. Often PCMs are modeled at such a high level that there is no clear link to work activities and processes. Many PCMs include only factors related to major phases of a program rather than more detailed work activities.

- They should be calibrated to your specific experience. Calibration is most likely the greatest challenge as most PCMs exercise predictions using data and modeled relationships based on data that comes from outside the organization. Often, these PCMs are initially based on aggregate industry data collected across many domains, life cycles, and so forth. As reinforced in Figure 18, calibration remains the most challenging and important modification to the traditional use of PCMs.

- They should be used to help compose processes. Traditionally, PCMs have been used to help forecast projects, bid on new work, and build initial project plans. However, Lockheed Martin found a greater value in using PCMs to help guide decisions about the composition of a process for a given project. In this endeavor, many cost analysts expert in the use of the PCMs found it more beneficial to engage project leaders with PCMs, thus enabling project decision making with regard to things like work activities.

- They should be used to support management what-if analyses. Although PCMs traditionally have supported what-if analysis during bids on new work and initial project planning, Lockheed Martin found it additionally useful to engage project leaders during key points of the project life cycle, thereby enabling what-if analysis during project execution as another means of decision support. This has been and continues to be a cultural change for Lockheed Martin project leaders.

- Actual data should be used to improve predictions at appropriate points and frequencies. The Lockheed Martin PCM experts began a data-collection activity to enable updated predictions during project execution and to improve the accuracy and precision of PCM predictions for future projects. The wider life-cycle uses of the PCMs have helped to build management sponsorship for the additional data collection and analysis.

- They should be probabilistic rather than deterministic and people should be trained to interpret them. Although many PCMs appear deterministic in structure without any notion of uncertainty or variation in factors and outcomes, Lockheed Martin used their PCMs in such a fashion that uncertainty in factors and ranges of outcomes were easily discernable.
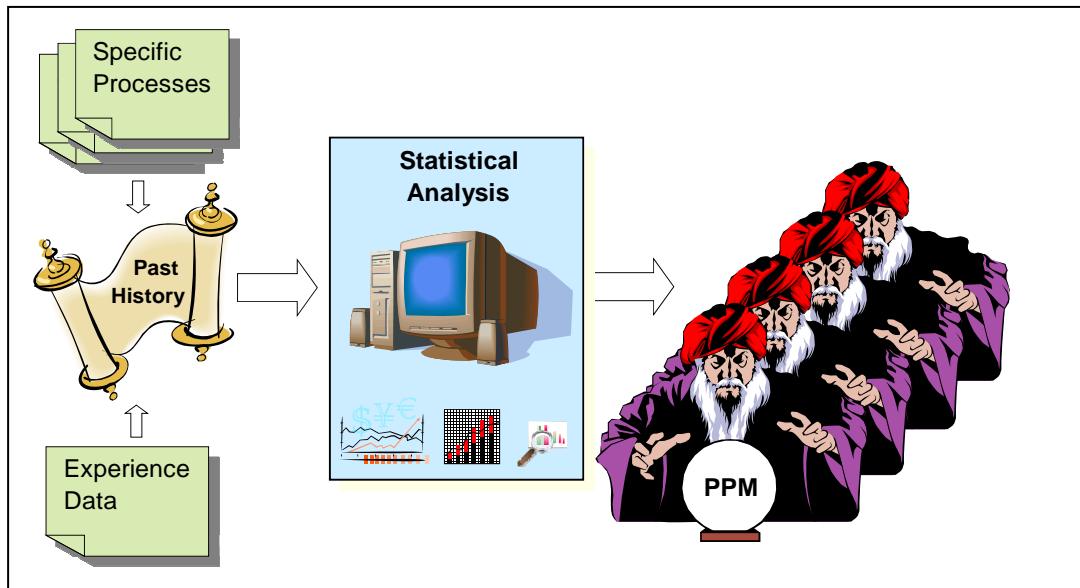
Figure 18: Parametric Cost Models as PPMs: Calibration is Critical

Lockheed Martin described a comprehensive and closed loop approach to the use of PCMs as process performance models.

Figure 19 illustrates the development, use, and calibration of the Lockheed Martin PCMs that serve as process performance models.
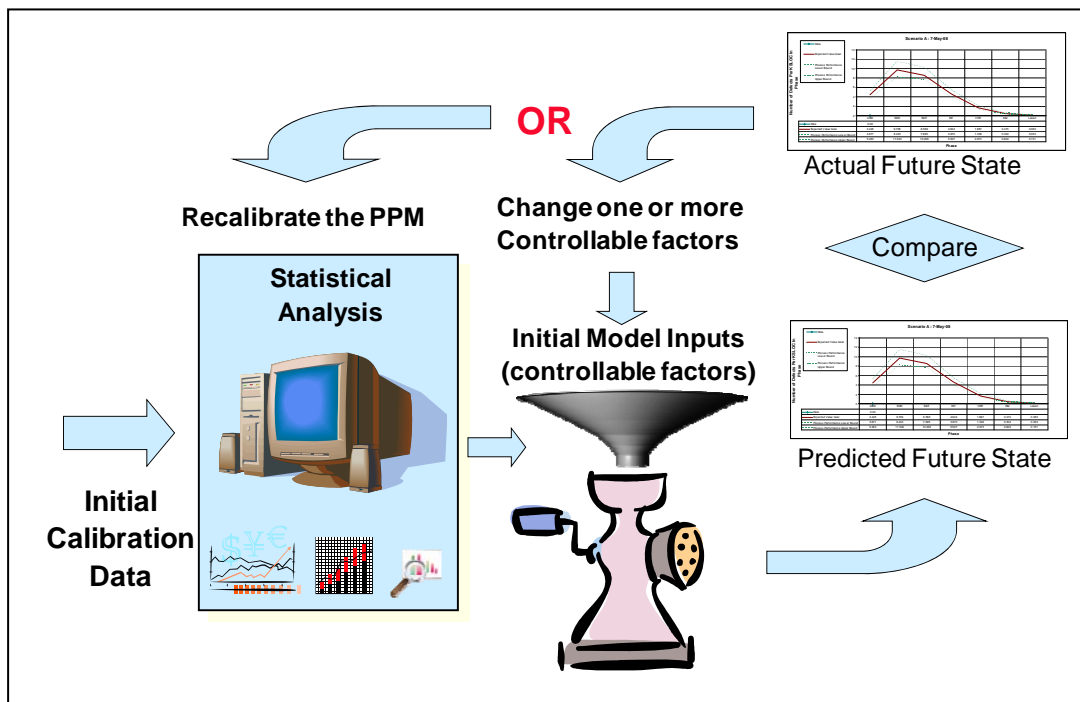
Figure 19: Development, Use, and Calibration of the Lockheed Martin PCMs

The COSYSMOR model, which refers to COSYSMO risk, estimates the probability distribution of systems engineering labor effort in addition to other outcomes. One outcome called effort risk

(i.e., the probability that actual effort will exceed target effort) may be predicted. COSYSMO and COSYSMOR are members of the COCOMO family of estimation models originally developed at the University of Southern California. Each of the COCOMO family of models, including COSYSMOR, is of the form:

$PH=A*(S^E)*D$, where:

PH = person hours
A = baseline unit effort (inverse of productivity)
S = size, number of equivalent requirements for COSYSMOR
E = exponent (indicative of economy of scale)
D = product of the cost drivers; A*D = the unit effort for the particular estimate

COSYSMOR provides a range of possible (estimated) values of resource (labor hours), NOT just the (single value) estimate. As shown in Table 6, uncertainty can be inputted in estimating parameters. The output will then provide "risk" for each value in a range of possible values of the form: effort risk = probability that actual value will exceed estimate.

The range of possible resource values and their risks can be used to support dialogue about how much risk to accept in a project, implementation alternatives, and so forth.

Table 6:   Uncertainty / Range Size of Parameter Values

	Uncertainty/ Range of Size Parameter Values								
	Low			Likely*			High		
	Easy	Nominal	Difficult	Easy	Nominal	Difficult	Easy	Nominal	Difficult
# of System Requirements	9	10	1	10	11	2	11	12	1
# of System Interfaces	2	10	3	2	11	4	2	13	5
# of Algorithms	3	9	2	4	10	3	5	11	7
# of Operational Scenarios	4	5	4	2	5	5	4	6	6

An example of the output from COSYSMOR is shown in Figure 20. This diagram depicts the time dimension in person hours on the x axis and risk probability on the y axis. In this example, the calculated risk probability for any given level of person hours can be identified.

Systems Engineering Person Hours Risk

In this example, the risk of exceeding 60,000 labor hours is slightly less than 20%.
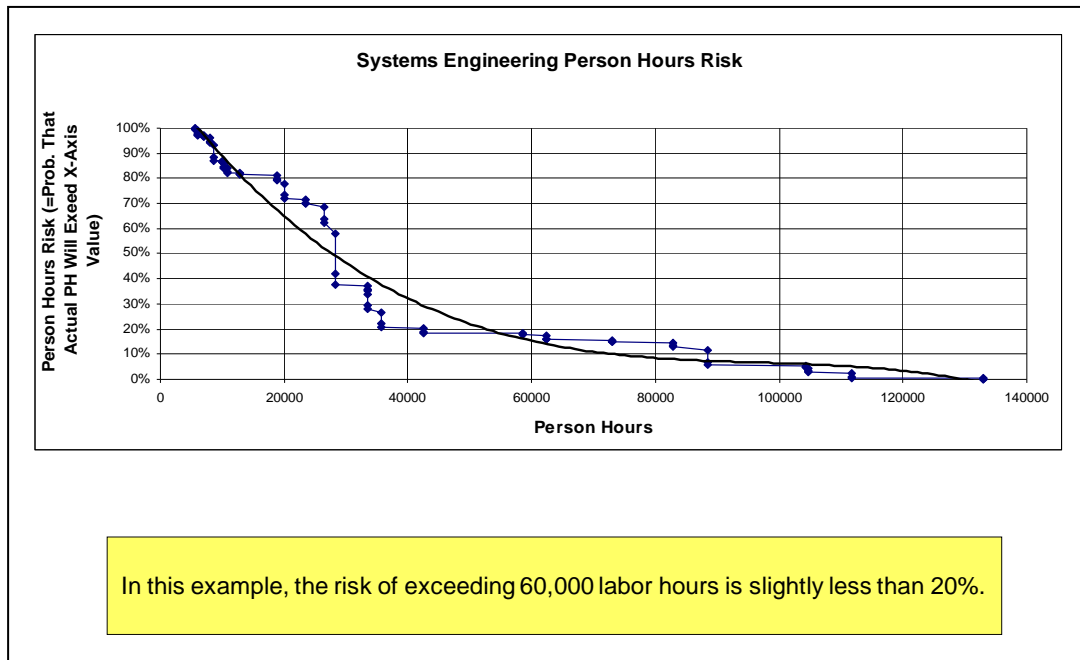
Figure 20: Output from COSYSMOR Showing Calculated Risk Probability per Level of Person Hours

Another example of the uncertainty modeling of an x factor is a cost driver referred to as the number of recursive levels in the design. The number of design levels related to the system of interest is depicted in the next table with associated systems engineering effort required for each level. In this situation, two controllable x factors are envisioned: 1) the number of recursive levels in the design, and 2) the degree of system engineering effort expected at each level.

Table 7: Design Levels and Associated Systems Engineering Effort

Attribute	Very Low	Low	Nominal	High	Very High
Number of levels	1	2	3 to 5	6 to 7	> 7
Required SE Effort	Focused on single product	Some vertical and horizontal coordination	More complex inter-dependencies coordination and trade-off analysis	Very complex inter-dependencies coordination and trade-off analysis	Extremely complex inter-dependencies coordination and trade-off analysis

These two controllable factors could be managed using what-if analysis of the two controllable factors to analyze the probability of meeting the project's goals and objectives. Depending on the results, potential management actions could be

• reduce the number of design levels required

- allocate the team's design assignments differently
- reduce the number of independent design teams to reduce coordination complexity

If the probability of meeting the goal is deemed acceptable, then immediate action is not necessary.

In summary, PCMs can be used as process performance models, but there are major distinctions between a traditional PCM and process performance models. Once those distinctions are well understood and additional project execution usage is embraced, PCMs can play an important role in process performance modeling. Finally, a set of PCM models may be useful to predict and manage performance during all phases of the development life cycle.

Process Performance Baseline for Northrop Grumman Mission Systems

Rick Hefner, Northrop Grumman Mission Systems (NGMS)

This presentation by Rick Hefner shared the different process performance baselines (defect injection, detection, removal, and escape) used to help predict productivity and defect data. An actual PPM (statistical, probabilistic, or simulation) was not part of the presentation. However, PPMs currently being used at NGMS were developed by the Office of Cost Estimation and Risk Analysis (OCERA), the same group responsible for maintaining the NGMS measurement repository called the Process Capability Database (PCDB). The PCDB dates back to the pioneering work of Barry Boehm in the 1970s, which led to the COCOMO parametric cost estimating tool. About 100 programs have contributed data to the PCDB throughout their program life cycles. A screening process is used to decide which of this data goes into the development of process performance models.

Within NGMS, the main business goals were to maintain, rather than improve, the average process performance. However, because such a wide variety of customers and program sizes and types existed, variation was an issue. Improvements were sought to reduce the variation, either through changes in the underlying organizational standard process or in the measurement system.

The quality of delivered products is difficult to use in goal setting because DoD cost-plus contracts often pay contractors to fix shoddy products. Thus, there is often a negative cost incentive to produce higher quality products, as long as the overall quality remains acceptable. As such, process performance models focused on productivity and defect density were the recent focus.

One of the key process performance models used was the Software Development Productivity Model. Based on project size and process tailoring choices, the model produced an expected productivity factor with a prediction interval. This model assisted projects in developing confident estimates of productivity and for making process choices during project planning. A similar process performance model was used for maintenance efforts.

Another key process performance model was the Defect Density Model, which provided projections of defect detection, defect injection, and defect leakage across the development life cycle from requirements through post delivery. There were multiple versions of this model to address different life cycles. Each version of the model showed actual with associated projections and confidence intervals across the phases. The confidence interval values were based on the organization's baseline data and were specific to the project parameters supplied by the user.

These models were primarily used in initial project planning and tailoring. There was limited business value in using these models to select or evaluate organizational innovations (i.e., in the Organizational Innovation and Deployment process area). Since NGMS was not seeking major changes in process performance, most improvements had little effect on overall project performance. In addition, because of the long timelines associated with DoD programs, changes to the organizational standard process description would take a long time to observe and be difficult to distinguish from other causes of variance (e.g., changes in DoD acquisition strategy, NGMS portfolio mix, experience, and educational level of the staff). Instead, classic Six Sigma techniques were used to analyze key subprocesses, identify localized improvements, and measure their effects.

Other process performance models are being built to address a myriad of factors outside of process and model their effects on program performance.

2.4 Statistical Methods

CAPRE Effort Estimation and Tracking Model Using Dummy Variable Linear Regression

David Webb, Hill Air Logistics Center

In this presentation, David Webb discussed his use of analysis of variance (ANOVA) and dummy variable regression to predict effort and schedule by phase using a variety of controllable process factors. This presentation included evidence of all the healthy ingredients for PPMs except for the connection of upstream and downstream activities. Webb acknowledged that several models could be used together to provide the upstream-downstream connection.

The Common Avionics Portable Reprogramming Equipment (CAPRE) project at Hill Air Logistics Center (ALC) replaces the obsolete, flight-line loader verifiers with COTS laptops using customized software and hardware. Webb described the following goals and objectives for this project:

- reduce costs by 80%
- reduce load times by up to 85%
- virtually eliminate repair costs (by using replacement instead of repair)
- introduce no new proprietary data or technology

The CAPRE project is intended for the following aircraft: C-130, C-141, C-17, C-5, A-10, HH-60, and MH-53. Future plans include implementation on the F-15 and B-52. CAPRE must be acceptance tested on an aircraft, which presents scheduling and staffing issues. Since the U.S. is in wartime, scheduling aircraft is challenging and must be done many months in advance. Most of the fixed-schedule tasks and milestones are well known and understood (e.g., duration of kickoff meetings, preliminary design reviews, critical design reviews, acceptance testing, and release/production). Additionally, an accurate estimate of effort for each subprocess in the Aircraft Adaptor Group (AAG) development process is required to complete the schedule and determine when aircraft will be required for acceptance testing. Any deviations from this effort estimate must be determined early enough to adjust available aircraft schedules. Therefore, the key model

required for CAPRE is an effort estimating and tracking model that can be exercised at project kickoff and updated at key milestones throughout the project.

The predicted outcomes (Ys) for the CAPRE model are as follows:

- total effort for the following subprocesses (to be used by the project manager):
    - planning
    - design
    - development (dev)
    - test
- schedule in days (to be used by the customer and project manager):
    - kickoff – PDR (planning)
    - PDR – CDR (design)
    - CDR – final bench test (dev)
    - FBT – developmental test
    - DT – operational test

Factors used in the CAPRE estimating model centered on the fact that each AAG project supports a unique, line-replaceable unit (LRU) computer. For 20 AAG projects, the following factors were collected:

- AAG type (x, nominal)

- number of operations (x, interval)

- number of LRU signals (x, interval)

- LRU age (x, ordinal)

- availability of test benches (x, ordinal)

- unique hardware (x, ordinal)

- documentation availability (x, nominal)

- LRU support (x, nominal)

- number of people (x, interval)

- actual total effort (Y, ratio)

Data were collected throughout the AAG project life cycles and at project closeout using a standard historical data worksheet. Effort data were collected using an online time accounting tool, while other data were gathered manually. Unfortunately, until the data analysis began, not all of the necessary data were identified, causing the historical data worksheets to be insufficient. For the 20 AAG projects researched, data not recorded on the historical data worksheet were gathered from the project documentation and subject matter experts. As a result, there was some risk of slightly inaccurate data (e.g., number of operations and signals) in the data set used for the CAPRE model.

For the CAPRE model, the mix of continuous and ordinal data required a one-way ANOVA using dummy variable regression. Using the SAS JMP fit model tool and examining the outputs for the Spearman's p correlation for each variable, the list was narrowed to four critical variables:

1. type

2. LRU age

3. availability of benches

4. unique hardware

Several modeling and data issues surfaced during the initial model development. Although the adjusted r squared and p values looked extraordinarily good, the prediction interval on the expression was very wide. Additionally, the equation produced negative predictions for some smaller values. From a data issue standpoint, all but three data sets were associated with AAG projects which required fewer than 5,000 hours of effort. This seemed to skew the data when including the few much larger AAG projects. As a result, the model development was repeated by limiting the AAG project data set to only those projects with <= 5k hours. The model was run again with the results shown in Figure 21.
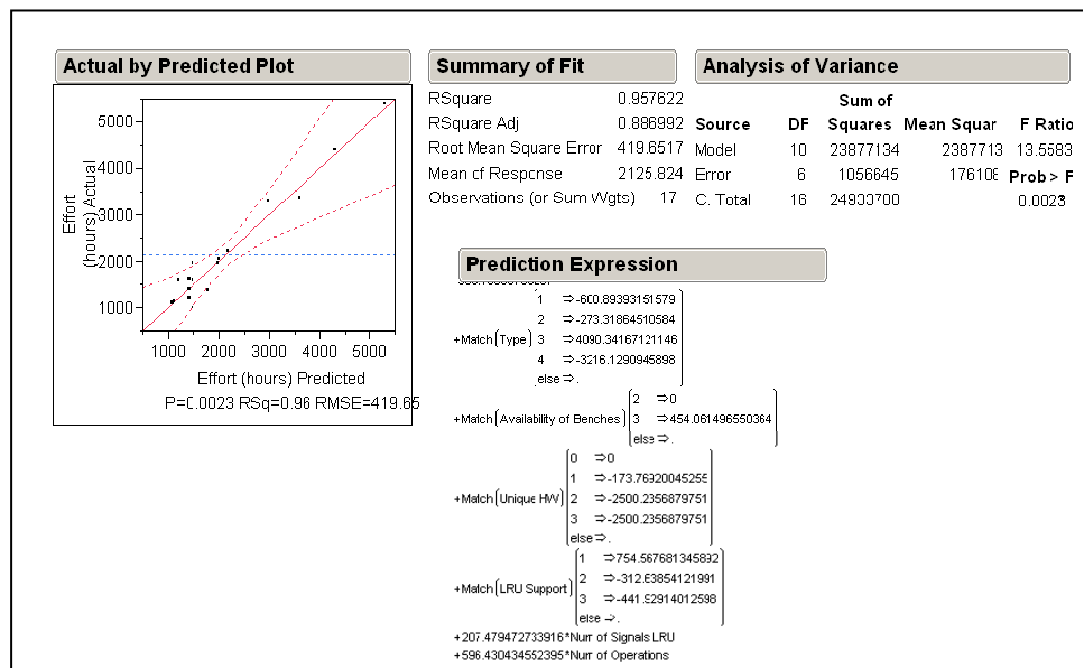
Figure 21: Results from the CAPRE Model

The newer model produced results more in line with experience and intuition. The prediction intervals were much more believable and helpful.

An example model of an Ethernet connection to an LRU in development with two operations and two signals was as follows:

• prediction = 4816 hours

• 95% UPI = 6883 hours

• 75% UPI = 5891 hours

Two ongoing challenges are described below.

1. The number of signals and operations is not well understood during project estimation. This value is needed as an input to the CAPRE model. Data analysis is being conducted to deter-

mine if there is a prediction model that may be employed to accurately predict the number of signals at the time this measurement is needed.

2. The model needs to be broken into subprocesses as shown in Figure 22.
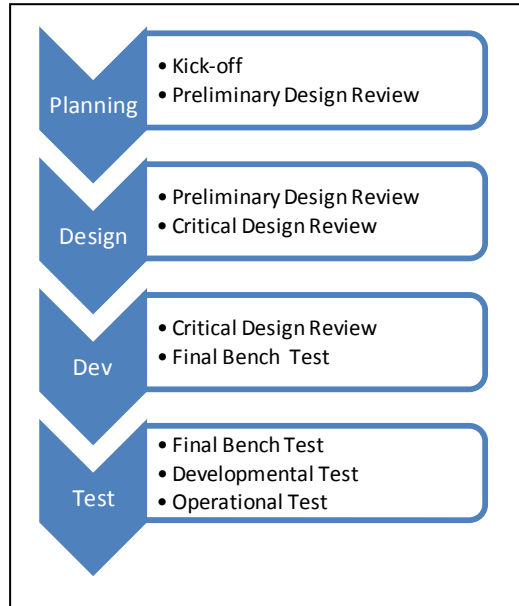
Figure 22: Subprocesses for the CAPRE Model

The model should then be run as each phase is completed, updating the prediction with the actual data most recently available.

In summary, the CAPRE project was able to create an effort prediction model using a few well-defined x factors. Prior to the CAPRE model, a model was in use that was not statistically based and included a number of factors without a statistical relationship to the prediction of effort. The CAPRE model can now be used to plan and track a project in support of the customer goals for aircraft availability. As discussed, several ongoing challenges exist and model refinement will be necessary to enable updated predictions at different subprocess events.

Customer Satisfaction Model

Lynn Penn, Lockheed Martin

In this presentation, Lynn Penn shared the development and use of the company's Customer Satisfaction Model (CSM). The model used multiple linear regression for predicting customer satisfaction (e.g., award fees) using a number of controllable program factors. This presentation demonstrated all the healthy ingredients of PPMs.

Systems engineering and technical assistance programs provide Lockheed Martin enterprise customers with rapid, in-depth technical advice, guidance, recommendations, and direction regarding the evolution, development, and operation of enterprise systems. In order to ensure a high level of customer satisfaction, a set of key factors needed to be identified and their impact understood.

The CSM was developed with the characteristics listed below.

- The program award fee (AF) was used as a measure of customer satisfaction.

- A functional relationship was defined between the AF and a set of controllable factors (i.e., $AF = k1*x1 + k2*x2 + \ldots$).

- Multiple linear regression analysis was performed using the available historical AF and factor data to determine a model and coefficients (ki), thereby enabling analysis leading to the optimum settings of the controllable factors and establishing the initial target baseline of customer satisfaction.

- The CSM model was designed to be calibrated periodically and updated based on the arrival of new data.

Table 8 depicts the typical award fee criteria and applicable indicators for one particular program type.

Table 8:   Award Fee Criteria and Indicators

Evaluation Category	Evaluation Percent	Controllable Process Factor	Applicable Factor Indicator
Management	25	Process Improvements Responsiveness	Audit Findings **(FR)** AI Closure Rate (**AI**)
Technical	30	Deliverable Quality	ERB Approvals (**AR**)
Cost	15	Cost Control	Cost Variances (**CV**)
Schedule	15	Action Item	AI Closure Rate **(AI)**
Other	15	Program Specific Items	Not used in model

This model, currently in the pilot phase, is intended for use before an award fee cycle to predict award fees based on the current values of the model's input factors. Where the prediction interval indicates a likely undesirable award fee, the factors that contribute most will be determined and causal analysis used to determine how the factors' values can be improved. Appropriate action plans will then be implemented. Since award fees are a significant part of the organization's profit structure, early warning signals of potential problems allow for a proactive organizational response instead of a reactive one that is too late to affect the award fees in a particular cycle.

The measures of the applicable factor indicators mentioned in Table 8 are as follows:

- **AR**: ERB approval rate – the percentage of artifacts approved by the program's internal engineering review board for transmittal to the customer during a given month

- **CV**: cost variance – the ratio of actual spending to planned spending during the award fee period

- **AI**: action item closure – the percentage of program management items due for closure during a given month that were actually closed during that month

- **FR**: audit findings rate – the number of audit findings divided by the number of process audits in a given month

The final model did not include the CV factor and did achieve statistically significant results, with an r-squared value of 0.72. The CSM will provide monthly updated prediction intervals for future AFs leading up to each six month AF cycle.

Tailoring Baselines and Models

Diane Mizukami Williams, Northrop Grumman Corporation

In her presentation, Diane Mizukami Williams described how she used multiple regression analysis to predict the tailoring hours needed for specific programs based on factors including the type of tailoring method. Almost all of the healthy ingredients for PPMs were evident in the presentation except for the connection of upstream to downstream activities and the enabling of projects to make mid-course corrections.

Northrop Grumman Mission Systems' baselines and models related to tailoring the internal development processes at project initiation were discussed. Northrop Grumman was focused on going beyond project and product baselines and models to create and use *process* baselines and models. Mizukami Williams shared baselines and models created by Northrop Grumman Mission Systems to help projects estimate their tailoring effort. The tailoring baselines and models were used to guide process improvement efforts aimed at continually improving the tailoring process. The presentation also discussed the goals for creating the tailoring baselines and models and showed actual tailoring data that was collected. Explanations of how the different baselines were established and led to a final tailoring model were provided.

The presentation emphasized the importance of having metrics in a thorough metrics program to address the three "P"s: project, product, and process. "Project" is commonly associated with CMMI maturity level 2, where programmatic metrics such as cost and schedule are collected. "Product" is commonly associated with CMMI maturity level 3, where quality metrics such as defects are collected. "Process" is commonly associated with CMMI maturity levels 4-5, where improvement metrics are collected, such as metrics to understand the effects of process changes.

Although it is important to have metrics to address all three, level 4-5 organizations often have baselines and models for project and product only. They have baselines and models for predicting project (i.e., cost and schedule) and product (i.e., defect) information, but what about process? At levels 4-5, organizations still fall back on "product" and re-use defect metrics for more sophisticated defect baselines and models. But baselines and models are also needed for process.

The Northrop Grumman Mission Systems Process Tailoring Model predicts the staff hours required to tailor the standard process for upcoming projects. The model was developed by the Northrop Grumman Mission Systems process management office (PMO). This model was created to provide guidance to projects on the tailoring effort needed and to help the PMO better understand the tailoring improvements. The model is routinely used during the start-up phase of new projects, several times a year. The PMO typically has one tailoring-related improvement per year, which can be significant. For example, one project spent 8,000 hours (over four years) to tailor the standard process. As part of this tailoring, each process area was assigned to a team of people who conducted several meetings to discuss each process step.

The business goals driving this model included

- understanding whether the tailoring hours were reasonable
- providing convincing data to management regarding the labor effort investment in tailoring
- providing a range of hours to help projects understand when they are spending too little or too much time on tailoring

The outcome predicted by the model was the estimated hours and a range of hours required for the tailoring activity.

The controllable factors used to make the prediction were

- method (Blank, pre-tailor, reused, or pre-populated)
- effort for tailoring (i.e., heads assigned to tailoring)

The non-controllable factors were

- development type (development, maintenance, services, or other)
- project type (software, hardware, or systems)
- project size (heads)

Additional information about the factors may be seen in Table 9.

Table 9: Factors in the Process Tailoring Model

Factor	Data Type	Controllable?	Baselines	Objective?	How Much Data	Factors Correlated	Other Useful Info
Development Type	Discrete	No	Same	Yes	227	Yes	
Project Type	Discrete	No	Min, Max, Mean, Std Dev	Yes	227	Yes	
Project Size	Continuous	No	Same	Yes	225	Yes	For analysis, size was converted to discrete (Micro, Small, Medium, Large)
Method	Discrete	Yes	Same	Yes	181	Yes	
Heads to Complete Tailoring	Continuous	Yes	Same	Yes	180	Yes	For analysis, heads were converted to discrete (1, 1 to 2, 2 to 3, 3 or more)

After 2002, Microsoft Excel spreadsheets were used for data collection since 47% of the 23,000 employees did not have easy access to the intranet. The spreadsheets contained macros that helped to ensure data completeness and integrity. All projects with more than 10 full-time-equivalent staff were required to complete the model input and prediction within 90 days of startup. A total of 184 projects contributed data toward the development of the model.

Minitab and Crystal Ball were used to develop the model. Minitab was used to analyze the raw data, test for differences, and confirm the significance of different factors. It also was used later to

create the regression equation for the final model. Crystal Ball was used to conduct Monte Carlo simulation of the model to show the sensitivity of the prediction based on changed behavior of one or more of the factors. As shown in the left graph of Figure 23, the tailoring hours were not normal (i.e., they failed the normality test). The tailoring hours were converted to lognormal data that passed the normality test, as shown in the right graph. Consequently, the model used lognormal data for the regression analysis.
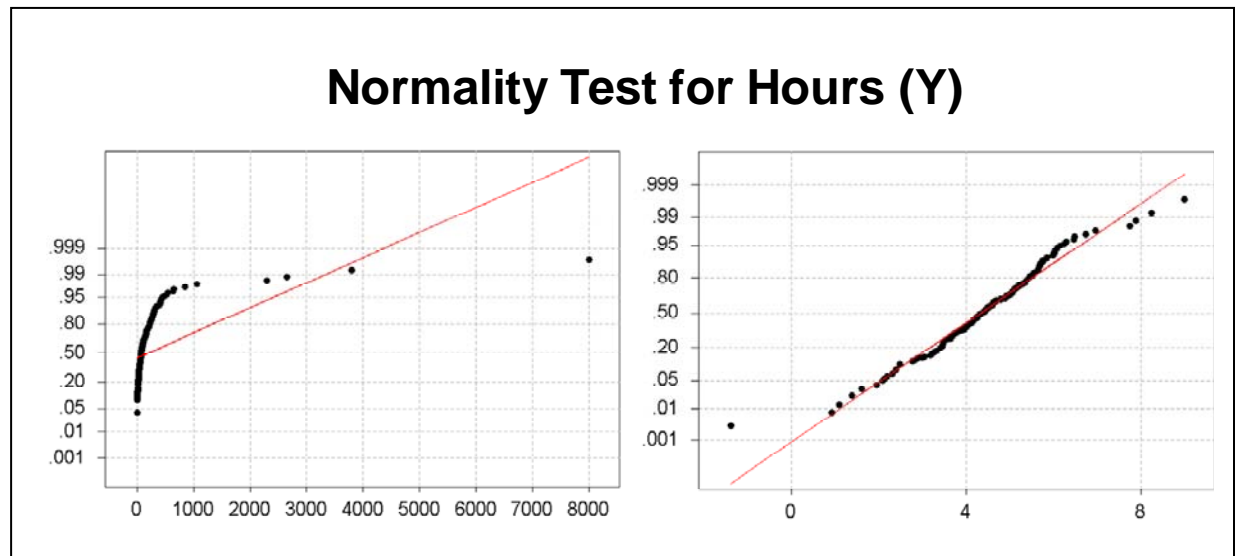
Figure 23:   Graphs Showing Normality Test for Hours

The initial data analysis sought evidence of process shifts in the amount of tailoring hours needed based on the tailoring method employed. As shown in Figure 24, the tailoring method improvements helped to reduce both hours and variation. Pre-populated tailoring appeared to be the most helpful (i.e., had the smallest median), but the Q3-Q1 range was high because that division made the tool more complicated by inserting many more things for the tailoring. Confirmation of this improvement is not definitive, though, as the 95% confidence intervals did overlap for all the methods (see Figure 24).
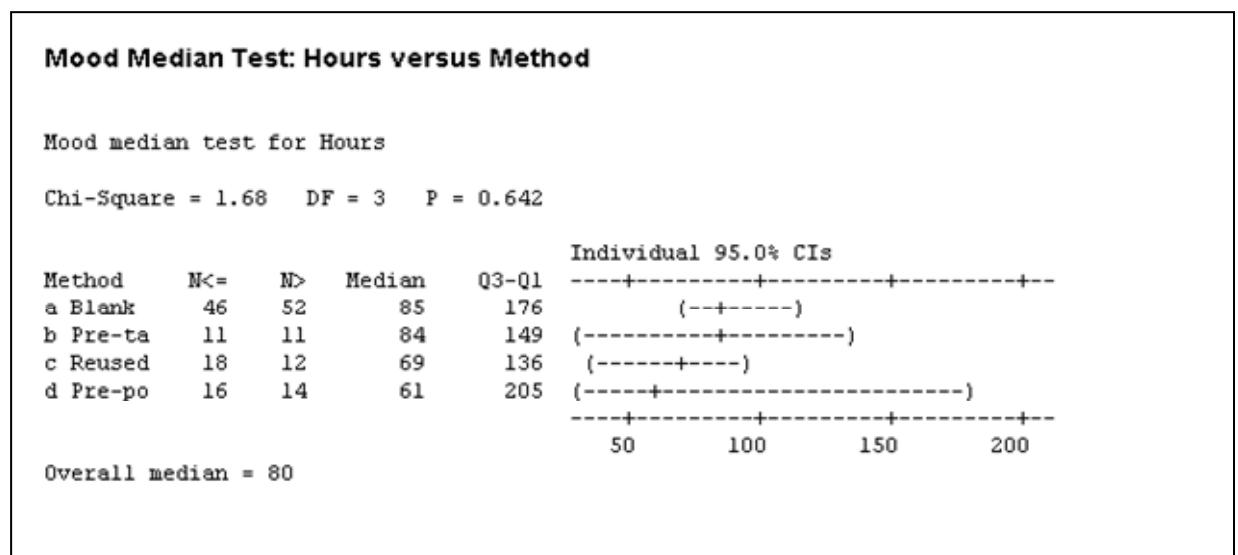
Figure 24: Mood Median Test

In addition to testing for differences of the median tailoring hours by method, a test for equal variance clearly shows the Blank method to be significantly statistically different (i.e., P-value = 0.001). See Figure 25. As a result of this outcome and the seeming difference of the mood median test, the tailoring method was included as a factor in the regression model.
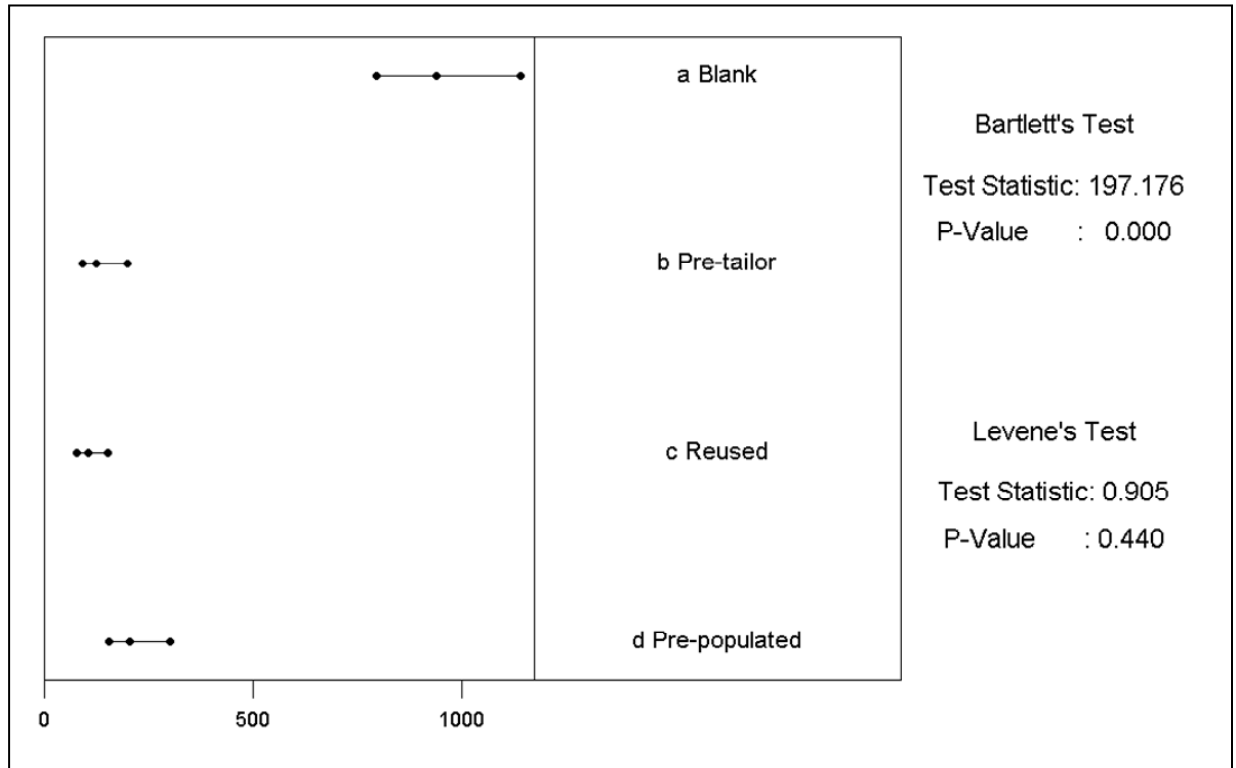
Figure 25: Tests for Equal Variance

The original Rough Estimation Model used only the historical baselines to provide a range of estimated hours, without benefit of a regression equation. Users would make selections from pull-down menus, shown in the screenshot in Figure 26. The model is called "rough" because the factors are not correlated with one another (i.e., some factors have a greater influence on the estimated hours). As such, a more accurate model using a regression equation was desired.

Tailoring Hours : Rough Estimation Model

		Q1	Q3	Median
1	**Project type is primarily:**			
	Development	46.5	230.0	97.0
2	**Development type is primarily:**			
	Software	45.1	227.5	98.5
3	**Project size:**			
	Small Project (< 20 FTEs)	24.5	104.6	49.0
4	**Method to complete the tailoring:**			
	c. Reused	22.0	153.5	69.0
5	**Heads to complete the tailoring:**			
	c. 2	45.0	161.0	84.0
	Estimated Hours: 37 to 80	**36.62**	**175.32**	**79.5**

Figure 26: Interface in Rough Estimation Model

The Regression Estimation Model, consisting of a regression equation, was developed using Minitab. Unlike the Rough Estimation Model that used pull-down menus (i.e., discrete data), this model used continuous data for project size and heads, which increased the accuracy and information value of the data. The Excel spreadsheet user interface for the Regression Estimation Model is shown in Figure 27. This model accounted for multiple project and development types and normal vs. lognormal tailoring data. This model was compared to the Rough Estimation Model to confirm the prediction was reasonable. The model took less than one day to build and takes less than one day to maintain and perform updates (done once annually).

Tailoring Hours : Regression Model

1 Project Type:

Development	Yes
Maintenance	Yes
Services	No
Other	No

2 Development Type:

Software	Yes
Hardware	Yes
Systems	Yes

3 Project Size (Number of FTE Heads): — 20

4 Number of heads to complete the tailoring: — 1

5 Method to complete the tailoring: — Reused

Estimated Hours: 38

Figure 27:   Interface in Regression Estimation Model

Monte Carlo simulation was performed using Crystal Ball to enable a depiction of the range and distribution of the predicted tailoring hours, which is more accurate than the Rough Estimation Model prediction of a single point estimate. Additionally, Crystal Ball provided fields during the analysis of the simulation results that allowed the user to enter percentages or confidence levels. The example in Figure 28 shows with 80% confidence that tailoring will take between 32.79 and 91.51 hours.
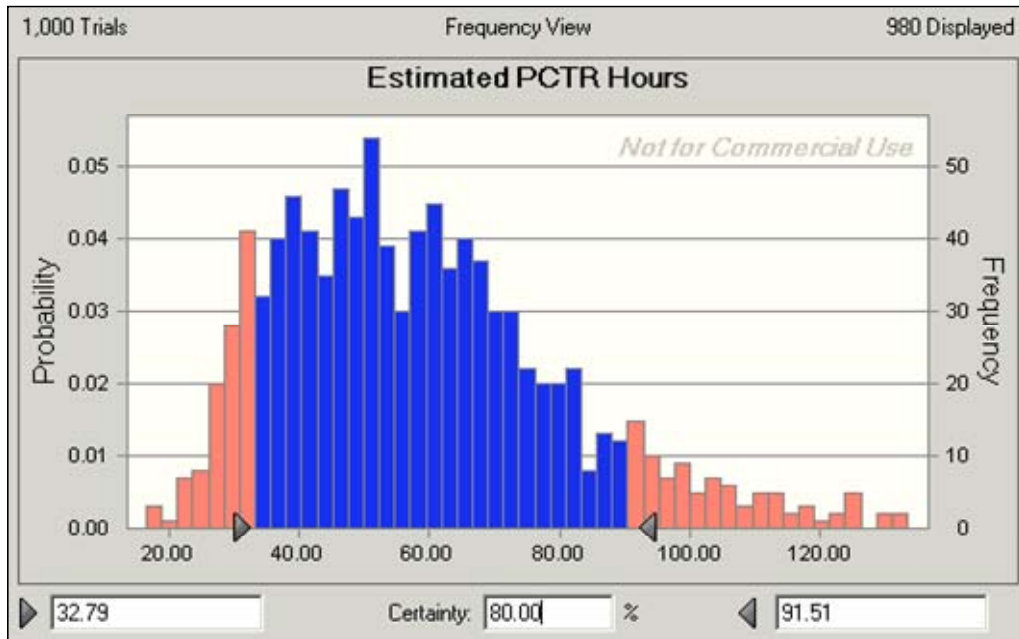
Figure 28: Crystal Ball Output of Regression Estimation Model

Key observations from the initial development and use of the model are listed below.

- The PMOs often thought the tailoring improvements would work, but data analysis showed they did not work.
- Using the model, projects understood that tailoring hours were greatly affected by the method they used for tailoring.
- During initial data analysis, the model developers discovered that more granular data would be needed. As a painful result, over 200 projects were contacted to provide more data.
- Model developers discovered that projects had interpreted tailoring hours differently, causing noise in the tailoring hours data.
- It was difficult to publicize the existence of the model across the entire organization so projects would use it.
- Small organizations with very similar projects or products have more consistent data than organizations with large and diverse projects. The challenge facing the Northrop Grumman Mission Systems organization was both size and diversity. The projects varied in nature from tracking blood for the Red Cross to creating missiles.
- The Monte Carlo simulation using Crystal Ball worked well, but the organization considered it impractical to purchase a global license for 23,000 people. Consequently, Microsoft Excel was used to create the probabilistic model.
- Arriving at the final regression equation for the model was difficult. Iterative attempts were required as some factors did not pan out while others did. Patience and detective work were required to arrive at a decent set of predictive, controllable factors.

Overall, the project to develop and implement the Tailoring Hours Model worked well and the model is now presented during the tailoring class. The organization has warmed up to the model

because a tool was created that automatically imports the tailoring data. Finally, it helped to have enough project data to create the model.

In summary, baselines and models must be of value to someone or they are a waste of time and money. They will fail if they are only created to satisfy CMMI maturity levels 4-5. As a result of the initial experience with the Tailoring Hours Model, existing baselines and models (e.g., peer reviews) are currently being expanded to add finer granularity of data with additional factors. Additionally, new baselines and models are planned for the new CMMI for Services, including a help desk.

Continuous Improvement for Cost of Quality
Mark Kelley, Esterline AVISTA

This presentation by Mark Kelley provided a brief glimpse of how simple linear regression was used to predict the differences in cost of quality based on process inputs from peer reviews and product attributes. Esterline AVISTA, a relatively small company employing approximately 150 people in Wisconsin and Minnesota, is a provider of safety-critical software development services in the aerospace, defense, and medical industries. Kelley described the company's development and use of two process performance models with a focus on cost of quality analysis. Based on relatively simple regression methods, the models were used to plan, monitor, and control the company's provision of testing, review, and defect correction services. The models were used by and benefited many stakeholders at AVISTA, including project teams, project leaders at different levels, company management, and their clients. While they focused on cost of quality, the models also were used for process improvement and to improve the quality of the delivered products by finding defects that needed to be fixed.

Estimation of verification effort often is based on average requirement complexity. Yet it often is impossible to fully understand exactly how complex the requirements will be until a project starts. For example, what kinds of automated tests will be necessary to verify that production code satisfies the requirements? AVISTA recognized that process performance models could help to accurately estimate the downstream work based on a better understanding of the requirement complexity and defined processes once project work started.

In fact, a major goal was to be able to calculate better estimates of project completion hours once a project started. Since there were strong links between the models' early and later project phase results, AVISTA was able to use early actual results to better estimate the effort required in later phases. The model predictions were used to establish, increase (and occasionally decrease) the staff count needed to do the work while meeting contractually agreed upon schedules within budget.

The first model used the baseline relationship between development hours per requirement and review hours per requirement to predict cost of quality, time per requirement, and total cost of quality hours for a project. This enabled projects to estimate rework hours as well. The second model helped the AVISTA engineers and their clients understand the cost of quality and the impact of quality on a project timeline. The model predicted the cost of quality hours per requirement and the total cost of quality hours for a project.

Several of the x factors in the data set used to fit the company's process performance models remained proprietary for the presentation. However, one notable controllable factor was the risk level AVISTA was willing to accept for a particular outcome being estimated. The work schedule was a notable uncontrollable x factor.

Other model factors included the following:

- development hours per requirement, used as a surrogate for the complexity of the requirements developed
- review hours per requirement, based on the assumption that the resolution of defects takes longer as the complexity of the tests increase
- defects found per requirement during review. Since more rework time generally is needed to fix the defects, more of them are discovered per requirement.
- project cost of quality hours per requirement (i.e., the sum of all time needed to fix the defects and review the resolutions of the defects)

The models and model predictions were based on straightforward regression techniques. The initial baselines used for both models yielded r^2 values over .9. Although the confidence intervals were quite narrow, prediction intervals were not used. When statistical outliers were found, they were researched and omitted from the baseline if special causes were determined. If the line of best fit intercepts the y axis at a negative number, the model could produce a spurious prediction of cost of quality for small values of x. Hence all predictions were limited to a minimum value of 0.

It took about 40 hours of data mining to discover the statistical relationships that proved to be useful for the predictive model. Considerably less time was necessary to tune the model for maintenance and new releases. The models were recalibrated and updated with new slopes, intercepts, and parameter estimates when the baseline data set was updated. While the basic statistical methods are reasonably straightforward, a rather elaborate estimation approach was used to determine what baseline elements and model specifications were to be used in managing the project in a particular instance. At the time of the presentation AVISTA also was considering how best to segment the data to improve the quality and pertinence of the model predictions.

The models were developed using a spreadsheet application and were peer reviewed by two levels of project managers before initial release. The team at the company looked for additions and modifications to the models when new baseline data sets were released. Both models were scheduled for periodic review to assess the impact of additional baseline data on the model predictions. At the time of the workshop presentation there had been one review, and the models were not changed.

The source data used for both models already existed in the company's automated system that tracks all work hours, reviews, and defects at a detailed task level. Over one year of historical data were used for the development of the two models.

Verifying data quality and integrity and maintaining the models is part of the company's standard process. This process includes the periodic extraction of source data using procedures stored in the database. The extracted data also are evaluated to verify that the correct source data was extracted.

While quality is important—and certainly vital in safety-critical applications—there are impacts to timelines and budgets that also need to be addressed. The models described in this presentation address all three. Simple visualizations appropriate for managers, such as pivot tables, are used to describe the model predictions and their implications. Using the models resulted in notable reductions in cost of quality and delivered defects. Comparable or better quality could be achieved at a savings within schedule. At times, AVISTA managers could pull staff off a project because they had confidence they could complete their service ahead of schedule and on target for quality. Thus they could reallocate staff to where they were needed more. Other benefits included the additional buy-in that resulted from more accurate estimation of project effort and schedule time, and there was a cultural shift within the organization toward making decisions based on historical data.

Training Improvement Through Competence Measurement & Analysis

Angel Liu, Motorola China

In this presentation, Angel Liu demonstrated how simple linear regression and ordinal logistic regression were used to predict manager competency and performance based on the deployment of improved processes for training, mentoring, and other professional development activities. Although some healthy ingredients for PPMs were not evident in the presentation (interim outcomes, upstream-downstream activity connections, and enabling projects to make mid-course corrections) they will be represented in future incremental work on the evolution of the modeling.

Liu reported on the use of process performance modeling at the Motorola China Center in the context of the Causal Analysis and Resolution (CAR) and Organizational Innovation and Deployment (OID) process areas. The work was aimed at improving the center's training processes.

The purpose of training at the center is to develop the skills and knowledge people need to perform their processes effectively and efficiently. The effectiveness of the organization's training processes is meant to directly influence process execution and therefore project success. Management believes that its first-line project leads are critical for the organization when business is growing rapidly, particularly because they are the direct interface with customers during the software production cycle. Similarly, they believe that development of their process quality specialists is critical for the organization to sustain its high maturity practices. Thus their emphasis is on how best to improve role-based training for the project leads and process quality specialists. The effectiveness of its training program is crucial for the center since it relies heavily on the use of training for skill development.

The organization's focus is on gaining a better quantitative understanding of project management competencies and their relationships to high maturity process quality and performance outcomes. Liu discussed competency measurement and evaluation and training needs analyses done at the center. The analyses were conducted using a variety of analytical methods to help identify the significant factors that influence the competence levels of the leads as well as the outcomes of their process performance. Actions were taken to improve processes for training design, mentoring, training package review, and related processes. The benefits of these improvements were verified through comparative analyses of performance outcomes.

The organization's standard monitoring and control activities showed a need for process improvement for both the project lead and process quality specialist roles. For example, a customer

satisfaction survey showed that customer satisfaction with project management was lower than that of other areas surveyed. Similarly, an audit finding showed training for the project leads was inadequate as they took on new project planning roles. Skill requirements for the process quality specialists also were increasing and changing, and expectations grew as the organization increased its need for higher levels of process capability. A Pareto analysis was done for 15 identified knowledge and skill areas from a project quality perspective. The data definitions were derived from the *Project Management Body of Knowledge (PMBOK)* [13] and augmented with areas derived from various SEI sources and the site's CMMI-based processes. Data were collected based on a self assessment of competence, a survey-based evaluation by supervisors, training records, training feedback, and an attribute measurement systems analysis (MSA).

Several analytical methods were used to establish the process performance baselines and models. These included hypothesis tests using, in varying contexts, analysis of variance, chi square, and Mann-Whitney U tests; least squares correlation and regression; and ordinal logistic regression with Goodman and Kruskal's Gamma measure of association. A generic competency model was built to identify improvement gaps and training needs. The model was used for skill self-evaluation, team competence evaluation, customer evaluation, and gap analysis. While focused on the Motorola project leads and process quality specialists, the staff at the center believed that the model could be customized by people in different roles in different organizations.

The performance patterns and improvement needs differed for the project leads and process quality specialists, so two projects were set up to improve competence separately. Since the initial analyses and their results differed considerably for the leads and specialists, there were very different layers in the two competency PPMs. Both analyses were focused on identifying skill gaps, but the nature of the skill requirements differed considerably by role. While the PMBOK was used as the major input for establishing the project lead skill structure, setting skill priorities was more important for the process quality specialists.

A measurement plan and operational measurement guidelines were defined before data collection in both instances. Goals included a 10% improvement in competence levels for both the project leads and quality specialists and a 30% reduction in the development cycle for training project leads. Improvement in the competence of the project leads was expected to improve customer satisfaction and reduce project management overhead and project management risk. Improvement in the competence of the process quality specialists was expected to improve the satisfaction of support projects, increase the specialists' abilities to detect quality issues, and improve their abilities in process consultation and quality control. Moreover, the expectation was that training effectiveness would be improved with clear and prioritized training needs based on predictions from a competence gap analysis.

Pertinent measures used in different contexts as either x or y factors for the project leaders included team management experience; team size managed; risk management knowledge and skill levels; causal analysis knowledge and skill; understanding data; process tailoring skill; and decision-making skill. Project leaders with less working experience and less experience managing smaller projects had more difficulty with process tailoring, understanding data, risk management, causal analysis, and decision making. Related findings included a need for better process tailoring guidelines for small-scale projects and a mentoring mechanism for new project leaders. It was clear that the design for project leader training did not fully match the center's skill training matrix. The length of the training cycle also was too long and had an adverse impact on attendance.

After process improvements were implemented, the cycle time for project leader development decreased by 50%, from 12 months to 6 months. Their competence evaluations improved on average by over 15% during the same time period, with a lower bound of almost 9% improvement against the baseline. Overhead for the software development projects was reduced significantly after deployment of the training process improvements while productivity improved by almost 13%.

Surveys based on team self assessments and competency evaluations by their project team internal customers were used as the primary method to collect data for the process quality specialist analysis. Key measures included skill scores, types of skills (alternative or mandatory), and weighted importance of the skills for the defined role. The skill evaluation assessed 110 identified skills in 22 skill areas and 5 higher level competency categories. The MSA analysis showed a positive relationship between the self-assessment of competence and internal customer satisfaction levels. However, there was no clear relationship between the importance scores from the project feedback and self assessment. The relationship was statistically significant, but a considerable amount of variation was left unexplained. Liu emphasized the need for further analysis to better understand the underlying relationships.

As with the project leads, the overall results for the process quality specialists were noteworthy. Higher competency levels were found among the specialists with quality and measurement knowledge rather than project and process management knowledge. Their software engineering and software development experience impacted their competence levels significantly, as did their pertinent measured skills and their ability to apply them. Their skills with respect to CAR increased markedly after the improvements in the training process were deployed. Their internal customer satisfaction ratings for "management satisfaction" also increased concomitantly. Their professionalism ratings improved by 15%, and the trend continued as time passed.

Program Staff Effectiveness Model
Brooke Eiche, Lockheed Martin Systems Integration

In her presentation Brooke Eiche discussed the development of the Program Staff Effectiveness Model, which used multiple variable regression to predict program success based on a variety of controllable and uncontrollable individual, team, and organizational factors. Although only a few of the healthy ingredients for PPMs were discussed in the presentation, Brooke indicated that ongoing work to fold in principal component analysis would enable coverage of the interim outcomes and several of the other healthy ingredients. The goal was to improve the productivity within a program by focusing on people issues such as current staff behaviors and characteristics.

Surveys were created and sent to program managers and technical leads regarding the issues while data was collected on the success of the programs in terms of cost, schedule, and technical performance. The data analysis focused on identifying the correlation of the staff behaviors and characteristics to the different success measures of the programs. Consolidated findings were reported to program management and a plan was developed to resurvey programs periodically to keep the prediction models updated.

The Program Staff Effectiveness Model is a multiple regression analysis model developed to predict the success of a program based on the determined little x factors shown in Table 10.

Table 10: Factors in Program Staff Effectiveness Model

Subprocess	Big Y High Level Business Objectives	Little y Smart Goals Subordinate Business Objectives	Big X High Level Process Measures	Little x Critical subprocess Measures	Process Performance Models & Baselines
Program Staff Effectiveness	Improve Productivity	Development plan; do regular succession planning for all critical positions; track education metrics; implement new hire on-boarding process; explore other models/techniques to better optimize the organization; expand participation in mentoring program by 50%	Program Staff Effectiveness	Staff turnover, Retention plan, Domain experience, Employee morale survey, Interruptions/task switching, On-time start, Communication amount, Skills gap analysis, Utilization of external resources, Staffing transition phase plan, PM relationship	Program Staff Effectiveness Process Performance Model

The factors were categorized as controllable, uncontrollable, or undefined:

- controllable factors
    - (Q1)[21] defined roles and responsibilities
    - (Q2) proper skills and training
    - transition phase plan
    - retention plan
    - staff turnover
    - project planning
    - (Q13) interruptions/task switching
- uncontrollable factors
    - team cohesion
    - employee morale
- it depends (undefined)
    - communication amount
    - PM relationship
    - domain (program) experience
    - job experience
    - utilization of external resources

---

[21] The Qs followed by numbers in parentheses refer to the question number from the survey.

Descriptions of the three questions that were statistically significant in the multiple regression analysis are provided below.

1. (Q1) Defined roles and responsibilities – the awareness employees have of their roles and responsibilities while working on a program or project.

2. (Q2) Proper skills and training –the skills and training necessary to drive program or task success and support the program's goals and needs.

3. (Q13) Interruptions / task switching – the amount of interruptions or interactions an employee has with another individual through phone calls or emails and the amount of switching an employee experiences between various job assignments. Table 11 provides more detail for each of the x factors used in the development of the process performance model.

Table 11: x Factors in the Process Performance Model

Input: Little x Factor Used in PPM – Interruptions/Task Switching	
Little x data type •**Continuous or discrete?** Discrete (on scale from 1 = many interruptions to 5 = few interruptions) •**Nominal, ordinal, interval, ratio:** nominal	**Controllable or uncontrollable:** controllable
Baselines of the factor (distribution, mean, std deviation): Mean = 2.9 Std Deviation = 1.29	**Whether the baselines are derived subjectively or objectively:** subjectively
How much historical data for x factor was available? Current survey results	**What x factors were correlated:** Defined roles and responsibilities = .357 Staff Turnover = .7 Job experience = .546
Any other useful operational definition info for x factor:	
Data Collection for Little x	
Data Items to be Collected: Interruptions/Task Switching question on survey: "Describe the amount of interruptions or employees switching roles." (on scale of 1 to 5)	**Collection Level**: By program lead engineer
Collection Periodicity and date range: Life of program, quarterly	**Collection Method: Automated or Manual?** Manual
Sampling approaches: Survey results from lead engineers	**Data Validation, Verification, and any threats to data quality / integrity:** Data verified by program lead engineer, Reflection on success of program, survey results may be biased based on opinions
Measurement system evaluation and Percentage of Measurement Error:	**Statistical control on any of the processes producing the data:** p-value of interruptions/task switching factor = .045

Input: Little x Factor Used in PPM – Proper Skills and Training	
Little x data type •**Continuous or discrete?** Discrete (on scale from 1 = very little skills and training to 5 = proper skills and training) •**Nominal, ordinal, interval, ratio:** Nominal	**Controllable or uncontrollable:** controllable
Baselines of the factor (distribution, mean, std deviation): Mean = 3.55 Standard Deviation = 1.19	**Whether the baselines are derived subjectively or objectively:** subjectively
How much historical data for x factor was available? Current survey results	**Whether the x factors were correlated:** Defined Roles and Responsibilities = .573 Succession Plan = .367 Retention Plan = .376 Staff Turnover = .419 Team Cohesion = .486 Own morale = .658 Others' morale = .633
Any other useful operational definition info for x factor:	
Data Collection for Little x	
Data Items to be Collected: Proper skills and training question on survey: "Employees in the program have the proper skills and training required to perform their jobs." (on scale of 1 to 5)	**Collection Level**: By program lead engineer
Collection Periodicity and date range: Life of program, quarterly	**Collection Method: Automated or Manual?** Manual
Sampling approaches: Survey results from lead engineers	**Data Validation, Verification, and any threats to data quality / integrity:** Data verified by program lead engineer, Reflection on success of program, survey results may be biased based on opinions
Measurement system evaluation and Percentage of Measurement Error:	**Statistical control on any of the processes producing the data:** p-value of defined roles/responsibilities = .053

Input: Little x Factor Used in PPM – Defined Roles and Responsibilities	
Little x data type •**Continuous or discrete?** Discrete (on scale from 1 = not defined to 5 = clearly defined) •**Nominal, ordinal, interval, ratio:** nominal	**Controllable or uncontrollable:** controllable
Baselines of the factor (distribution, mean, std deviation): Mean = 4.3 Std Deviation = .979	**Whether the baselines are derived subjectively or objectively:** subjectively
How much historical data for x factor was available? Current survey results	**What x factors were correlated:** Communication Amount Team Cohesion Employee Morale
Any other useful operational definition info for x factor:	
Data Collection for Little x	
Data Items to be Collected: Defined roles/responsibilities question on survey: "The roles and job descriptions of employees in the program are clearly defined and identified to perform their jobs." (on scale of 1 to 5)	**Collection Level:** By program lead engineer
Collection Periodicity and date range: Life of program, quarterly	**Collection Method: Automated or Manual?** Manual
Sampling approaches: Survey results from lead engineers	**Data Validation, Verification, and any threats to data quality / integrity:** Data verified by program lead engineer, Reflection on success of program, survey results may be biased based on opinions
Measurement system evaluation and Percentage of Measurement Error:	**Statistical control on any of the processes producing the data:** p-value of defined roles/responsibilities = .081

The functional structure of the multiple regression model in Figure 29 is provided to highlight the positive and negative effects the factors have on the predicted success measure.

Success Rank= .408 + .32*Q1 - .25*Q13 + .29*Q2

Where   Q1 = defined roles and responsibilities

            Q13 = interruptions/task switching

            Q2 = proper skills and training

Figure 29: Functional Structure of Multiple Regression Model

The overall model was significant with a p-value of .009.

The initial approach used to develop this model will be repeated quarterly for feedback to program management and updates to the model. Additionally, program management training will be

supplemented with results from the modeling, including reinforcement training on the use of the model during program execution. As specific programs use the model to make predictions, leadership will be briefed and the model and analysis used to support causal analysis to improve the predicted success level by influencing the behavior of one or more of the little x factors.

Productivity is the initial focus for program management when using this model. Achieving earlier awareness of potential problems will enable a proactive response rather than the traditional reactive response that often comes too late to affect productivity.

Quality PPM Development for F-16 Operational Flight Programs

Dan Bennett, Rushby Craig, and Kevin Tjoland, Ogden Air Logistics Center, 309th Software Maintenance Group, Hill AFB, UT

This presentation by Dan Bennett demonstrated the use of multiple linear regression on data from the peer review process to predict the number of action items and escaped defects. His presentation provided evidence of all the healthy ingredients of PPMs. The 309th Software Maintenance Group is applying Six Sigma modeling and prediction techniques taught recently by the SEI. Dan Bennett described the Quality Process Performance Model created for use during the development project for the F-16 operational flight programs.

The initial target audience for this model was the lead of the operational flight program engineering team, within the fire control computer subsystem, who would use the model during the coding phase of development to anticipate and prevent downstream issues. The primary business goals for using the model were to

- improve quality
- improve cost performance
- improve defect detection ratio
- enable the earlier detection of defects

To determine the business goals, the Hill Air Logistics Center (ALC) improvement team surveyed senior management about the top 10 issues that they believed were preventing quality products from being delivered on time. From this activity, the code peer review process was identified as a key subprocess driving on-time delivery. The improvement team, along with domain experts from the project, brainstormed the x factors that might explain the performance of the peer review process. The two outcome measures having the greatest importance to the lead of the development team, the development team itself, and the technical project manager were operationally defined as

- number of major action items (continuous data type)
- number of escaped defects (continuous data type)

The x factors used in the model to predict the number of major action items were defined as

- # of reconvenes (number in 1-7 range, continuous, uncontrollable)
- experience of reviewers (proficiency rating, continuous, controllable)
- experience of author (proficiency rating, continuous, controllable)

- estimated effort (hours, continuous, uncontrollable)
- reviewer preparation (hours, continuous, controllable)
- # of participants (number in 3-20 range, continuous, controllable)

As the analysis and modeling began, the data were segmented for analysis and review. A homogeneous segment ranging from 20-3000 SLOC was used for the analysis presented at the workshop.

The factors used to predict the number of escaped defects were

- # of participants (number in 3-20 range, continuous, controllable)
- participant experience (proficiency rating, continuous, controllable)
- # of major action items (number, continuous, uncontrollable)
- # of minor action items (number, continuous, uncontrollable)
- delta literal memory (# of bytes, continuous, uncontrollable)
- delta data memory (# of bytes, continuous, uncontrollable)

Again, the same data segmentation was used to conduct the analysis and modeling to predict the number of escaped defects.

The data collection consisted primarily of data entered by the configuration management group into a tool that stores the data. As data mining was performed to extract the data, some questions about data quality arose that warranted further investigation. For example,

- Were preparation hours being reported accurately?
- Was statistical control used?
- Was an MSE analysis done?

The Hill ALC team pursued the use of iterative multiple regression beginning with 12 factors and narrowed them down to 6 critical, statistically significant factors previously discussed. The resulting regression model achieved an adjusted r-squared value of 75% with x factor p-values below 0.05. The regression analysis was performed in both JMP and Minitab, depending on the preference of the team member involved.

The challenges experienced in this modeling activity included

- identifying the correct model to develop
- identifying the correct outcome to predict
- identifying the correct x factors to model
- obtaining sufficient and meaningful data
- verifying data quality and integrity
- analyzing the model results
- communicating results and conclusions to stakeholders
- maintaining the models
- building and maintaining management sponsorship for modeling
- documenting evidence of use of the models

In summary, the organization and improvement team benefited from seeing the commonality between subsystems and planned to use what they learned from applying the model with the fire control computer subsystem on other subsystems. The process performance modeling development and usage is being piloted in several places within the ALC since initial experience has demonstrated the potential for significant insight and business value. However, some questions remain about where to best apply this modeling technique for significant return on investment. Additionally, concerns remain about the cultural change needed for the organization to develop, use, and benefit from the process performance model.

2.5   Modeling with TSP

Uses of Monte Carlo Simulation for TSP Teams

David Webb, Hill Air Logistics Center; David Tuma, Tuma Solutions; Jim Van Buren, Draper Laboratory, and Robert Stoddard, Software Engineering Institute

In this presentation, David Webb shared information about the use of Monte Carlo simulation to predict quality and schedule at the individual and team level. The presentation included evidence and discussion of most of the healthy ingredients of process performance models (PPMs), with the exception of controllable factors, connecting upstream with downstream activities, and enabling projects to achieve mid course corrections using the factors in the model. Webb and his colleagues plan to extend this work in the future by including these ingredients.

Webb described work done at Hill Air Logistics Center using PPMs to enhance existing work based on the Team Software Process (TSP). Large, consistently measured historical data sets already existed as a result of Hill's TSP initiative, as did quality and schedule models with a focus on practices that map well to CMMI Causal Analysis and Resolution (CAR), Organizational Process Performance (OPP), and Quantitative Project Management (QPM). For example, they had been tracking earned value, which was particularly useful since they were not having problems with product quality. However, their TSP models did not analyze variability of the historical data for planning purposes since traditional earned value techniques calculate single projections.

As a result of their focus on CMMI high maturity practices (and process performance modeling in particular) the engineers at Hill analyzed variability to generate uncertainty ranges for their projections, recognizing that wide variation intervals meant that point projections are a "stab in the dark" and should not be used for planning purposes. They used in-progress project data and analyzed the variability in them to refine and update their projections, using projection ranges to enhance their progress tracking. They reported that understanding variability helped them better manage and improve their processes and track them at the level of subprocesses. They analyzed several years' worth of plan-vs.-actual earned value data. A lognormal distribution worked well for predicting both task and schedule planning error.

Webb emphasized the use of Monte Carlo simulation models to better understand process outcomes. Monte Carlo has been particularly useful for their purposes because it promotes an understanding of the distributions inherent in all data and the estimates based on them. Examples described included an analysis to determine whether a trained but inexperienced Personal Software Process (PSP) engineer would be more or less likely to produce a low defect product than an experienced engineer not trained in PSP. Another analysis used standard TSP/PSP task and schedule plans for individual staff members in a parametric earned value approach with probability distri-

butions for task cost and calendar time. The data were used in a Monte Carlo model to estimate variability in the completion time for the team schedule.

The presentation also included a brief overview of Monte Carlo analysis, including its statistical limitations; reviewed the key TSP process performance models (project planning and quality management) from a Monte Carlo perspective; and discussed the historical data available to a TSP project that can be used to support Monte Carlo analysis of TSP process performance models.

Their experience thus far has convinced Hill analysts that Monte Carlo simulation holds great promise in filling the gap between TSP and CMMI high maturity practice. Within the TSP planning and plan-monitoring framework, this technique can be applied during the launch process and during project execution. An anticipated next step at the time of the presentation was using PPMs for defect management to better understand defect injection and phase yield distributions.

Process Performance Models and PSP/TSP

Kathy Smith, Hewlett Packard

In this presentation Kathy Smith described the use of traditional TSP quality profiles and other in-process, phase-based data to predict schedule and final quality. The primary healthy ingredients evident in the presentation included the prediction of interim and final outcomes, use of controllable factors, and the enabling of projects to make mid-course corrections.

Hewlett Packard (HP) organizations develop performance baselines and models based on the type of work they perform and their business needs. Performance models are developed by local organizational quality engineers and used by the projects as appropriate, at start-up and throughout the life of the projects. While organizations strive to develop models that are predictive and provide value for the projects, challenges continue, particularly in organizations where the focus is on supporting legacy systems and applications management. The most notable challenges center on data issues such as data quality, integrity, and granularity. These issues must often be resolved before continuing with the statistical analysis necessary to support the baselines and models.

HP has recently begun implementation of a PSP/TSP program. This program was initiated by the global strategic capability program as a productivity initiative not based on an engineering process group (EPG) or CMMI initiative. The executive seminar for TSP was held in December 2007, and the first five HP[22] TSP coaches were trained from January to May 2008. The first TSP pilot began in June 2008.

While the PSP/TSP program was not initiated as part of a process improvement or CMMI, the strong relationship between CMMI and PSP/TSP quickly became apparent. PSP focuses on adherence to processes, planning, estimating, and the use of measurement data at the individual level. A team applying TSP consolidates selected individual data to measure team quality, productivity, and progress. TSP also incorporates an emphasis on data quality, process adherence, and use of data to manage the project. When teams apply TSP, project goals are determined and confirmed during the project launch based on the sponsor's objectives. Team estimates and planned accomplishments are based on historical quality and productivity data from team members. TSP projects develop a detailed quality plan that is updated and reviewed weekly by the TSP project team.

---

[22]   The organizational units were part of Electronic Data Systems (EDS) at the time that the training took place.

More importantly, both the individuals and teams use this plan and associated data to understand their progress, quality, and productivity and to set goals for continuous improvement. These TSP activities provide a solid foundation for achieving CMMI maturity levels 4 and 5. Figure 30 and Figure 31 illustrate the nature of the measures and information used by the TSP teams.

- LOC/hour
- % Reuse (% of total LOC)
- % New Reuse (% of N&C LOC)
- **Percent Defect Free (PDF)**
- In compile
- In unit test
- In build and integration
- In system test
- **Defect/page**
- Requirements inspection
- HLD inspection
- **Defects/KLOC**
- DLD review
- DLD inspection
- Code review
- Compile
- Code inspection
- Unit test
- Build and integration
- System test
- Total development

- **Defect Ratios**
- Code review/Compile
- DLD review/Unit test
- **Development time ratios (%)**
- Requirements inspection/Requirements
- HLD inspection/HLD
- DLD/code
- DLD review/DLD
- Code review/code
- **Appraisal /Failure Rates**
- **Review rates**
- DLD lines/hour
- Code LOC/hour
- **Inspection rates**
- Requirement pages/hour
- HLD pages/hour
- DLD lines/hour
- Code LOC/hour

Figure 30: TSP Quality Plan One

- **Defect-injection Rates (Defects/Hr.)**
- Requirements
- HLD
- DLD
- Code
- Compile
- Unit test
- Build and integration
- System test
- **Defect-removal Rates (Defects/Hr.)**
- Requirements inspection
- HLD inspection
- DLD review
- DLD inspection
- Code review
- Compile
- Code inspection
- Unit test
- Build and integration
- System test

- **Phase Yields**
- Requirements inspection
- HLD inspection
- DLD review
- Test development
- DLD inspection
- Code review
- Compile
- Code inspection
- Unit test
- Build and integration
- System test
- **Process Yields**
- % before compile
- % before unit test
- % before build and integration
- % before system test
- % before system delivery

Figure 31: TSP Quality Plan Two

TSP uses quality profiles to predict the ability of the team to meet project objectives. Although not statistically based, the quality profiles are intended to provide an initial capability that aligns to the definition of process performance models as defined in the CMMI for Development Version 1.2: Glossary Definition of Process Performance Model:

> *"A description of the relationships among attributes of a process and its work products that are developed from historical process-performance data and calibrated using collected process and product measures from the project and that are used to predict results to be achieved by following a process."*

The project team uses TSP quality profiles to predict which software components (programs) are likely to have quality problems by viewing Kiviat diagrams that display key factors associated with each component (see Figure 32). By comparing the diagrams to one another, it is possible to identify components that are likely to have problems. These diagrams provide similar information to the orthogonal defect classification (ODC) footprints developed by IBM and may be used by the project team to anticipate future project production or execution issues. The project team can also use the quality profiles to identify which processes are causing the problems and improve them. Finally, the testing team can use quality profiles to plan for system testing, similar to how ODC footprints were used to help software test planning.

A quality profile is composed of the following x factors:

- design/code time
- design review time (formula)
- code review time (formula)
- compile defects/KLOC (formula)
- unit test defects/KLOC (formula)

These factors are considered controllable. They are based on continuous, ratio data, and are updated incrementally. The amount of available historical data will increase as more TSP projects are completed at HP. The HP TSP pilot teams were able to use initial TSP data provided by the SEI. Figure 32 depicts examples of several quality profiles based on those data.[23]

---

[23] See http://www.sei.cmu.edu/library/abstracts/whitepapers/qualityprofile.cfm for more information about quality profiles.

Figure 32: Quality Profiles

While the quality profiles do not meet all the expectations for CMMI process performance models, they have the advantage of being an integral part of all TSP projects, and they are actively used by TSP teams to measure interim progress toward meeting project objectives. Potential improvements to the PSP/TSP analysis could include the use of factors that indicate variation and the shape of the distribution of the data instead of a simple computed average. Additionally, a next step could be to establish statistical regression models using the x factors in the quality profile to predict project success measures. Two notes of caution when creating baselines and models based on PSP/TSP are described below.

1. Ensure the data quality and integrity will support statistical analysis and regression modeling.
2. Focus on the value-added for the users of this information, not just meeting CMMI maturity level 4 requirements (e.g., ensure compelling business value exists for additional statistical techniques that are considered).

Process Performance Baselines and Models in PSP and TSP

Yoshihiro Akiyama, Kyushu Institute of Technology & Next Process Institute

In this presentation, Yoshihiro Akiyama described the use of traditional TSP measures of quality, including the product quality index, to predict final schedule performance. Akiyama also shared modifications using uncertainty distributions for some of the baselines to help judge future performance. The primary healthy ingredients evident in the presentation were the characterization of variation and uncertainty in the factors and the pursuit of mid-course corrections by projects when the distributions of the outcomes did not look acceptable. Akiyama also discussed how TSP processes can be refined to support statistical modeling of interim and final project outcomes as a function of controllable factors, what-if analysis, and connecting upstream and downstream activi-

ties. In addition, he discussed several excellent TSP practices to ensure quality and commitment that do not map directly to statistically predictive modeling.

After a quick overview of the measurement frameworks of PSP and TSP, Akiyami described a few basic process performance models that have been used in TSP implementations. He discussed how the baselines and models were defined and used for planning and control purposes to manage process performance objectives throughout the project life cycle. He also discussed two examples that the TSP process covers but that are not covered in the CMMI models. He first discussed how process performance models used in TSP planning are reevaluated for their accuracy early in every project phase. Next he discussed how TSP concurrently emphasizes meeting the project delivery dates when managing the quality performance of subprocesses associated with specific engineering artifacts such as product components.

Akiyama emphasized that high maturity practices must be owned intellectually by management and engineers at all levels of the organization, with clear bi-directional communication among them. He also emphasized that high maturity practices must provide tailorable frameworks based on data to be used for project planning and control, with a focus on performance-based statistical thinking. All of these criteria are of the essence of PSP/TSP as well as high maturity measurement and analysis.

Much of the presentation was used to map concepts and terminology in PSP/TSP to CMMI-based PPMs and PPBs. For example, planned estimates and actual values are contained in PPBs. Subprocesses are built into individual and team data as a matter of course, with a focus on planning related to schedule, quality, throughput, and so on. Those data in turn can be used to monitor status and be used for prediction in PPMs. Akiyama also noted that PSP/TSP estimation now includes prediction intervals, as influenced by the PPM healthy ingredients. On a related note, non-attribution at the individual level is enforced with existing TSP tools. While project level data are derived from individual data, only team-level data are reported to management.

Akiyama also spoke at some length about the use in PSP/TSP of proxy-based PROBE size estimation, which can be useful since very little information is available for estimating during the early stages of many projects. Since the project size estimates are based on aggregates, variation and errors of individual-level data can be offset.

The presentation also included several examples of indicators and analytical methods used for estimating, monitoring, and controlling project outcomes. Individual and consolidated earned value curves commonly are used in TSP. Although there was little explicit discussion of what-if predictive modeling, it too was implied by the discussion of the importance of tailorability. Along with control charts and other more common graphical representations, the examples included several spider charts to show the interrelationships among the components of a TSP process quality index (PQI) with quality profiles before and after test. Such indicators can be quite useful for communicating with management and engineers who are not expert data analysts.

While promoting such understanding has been a major goal of PSP/TSP, the emphasis also has been on doing so by helping engineers and managers come to appreciate the value of data and statistical thinking. To that end, Akiyama noted that lognormal distributions have been used for size measures, where each size category has its own distribution. An example of the relationship between PQI and post-development defects suggested a need for further regression-based investi-

gation to account for the wide variation in defects at the lower PQI values (e.g., by including additional variables in the analysis, possibly at the individual level).

Akiyama discussed a broad picture of how subprocesses and measureable quantities are related to each other, which is not described fully in CMMI models. He proposed a "measure–subprocess" relationship, such that for every direct or derived measure, there exists a set of subprocesses that contribute to that measure. He asserted that it is important to take an over-arching approach in defining the measurement and analysis framework on all possible measures and subprocesses supporting those measures. In other words, the measures need to be based on the voice of the process. When one of the measures shows fluctuation, the related subprocesses should be examined and controlled in order to eliminate the undesirable variation or for replanning.

3   Other Workshop Sessions

3.1   Special Session: Improving Estimates of Prediction Error

Mike Konrad, Software Engineering Institute

How best to represent variability and understand the predicted range or variation of model out-comes remains difficult, even for those who are not new to CMMI-based process performance modeling. Mike Konrad gave a short presentation describing the implications on the topic of some recent statistical literature.[24] His focus was on bootstrap, cross validation, and jackknife resam-pling statistical methods.

There are many resampling methods in statistics. They are used to estimate the accuracy (i.e., pre-cision) of descriptive statistics that are calculated based on samples drawn from larger populations about which generalizations are being made (e.g., medians, variances, or percentiles). Bootstrap methods repeatedly draw samples randomly with replacement from a set of data points. Jackknife methods work by using subsets of the available data. Cross validation methods also use random subsets. Like bootstrap methods, cross validation methods are used for model validation.

Konrad began by emphasizing that estimating, measuring, or otherwise assessing the accuracy of model predictions is one of the most critical problems that needs to be solved in any kind of prob-abilistic or statistical modeling, whether the subject matter is product, process, or the performance of an organization, project, team, or individual. Another important consideration is how much variability there is likely to be around a model's predictive estimates and the values that are ob-served subsequently.

Prediction error can be characterized for both continuous and discrete data. Few assumptions about population distributions are made for resampling methods. Konrad also pointed out that graphical displays can be used to make the results easier to understand for those using the model results. PPM examples could include a regression model predicting error density of a product component entering system test or various classification models used for prediction.

Bootstrap methods effectively treat a larger initial sample (sampling frame) as the population for repeated resampling. Repeated samples of the same size are drawn from the original sample. How many times must you resample? Efron characterizes the solution from both a layperson's and a mathematician's perspective [12].

Cross validation works by comparing test samples (validating sets) with a separate training sam-ple (training set). The test samples are subsets of the original data set that are excluded from the training sample. The model is fit initially to the training sample and used to predict for the test samples. An average of the predictions across the test samples is used as an overall measure of prediction accuracy. If the model is built without using any of the test sample data, then an un-biased estimate of prediction error is more likely.

---

[24] The presentation was based primarily on [14]. Further explication and the formulas that Konrad described in the presentation can be found there. A good overview of resampling statistics also can be found on at http://en.wikipedia.org/wiki/Resampling_(statistics).

Jackknife works by re-computing the estimate for each statistic by leaving out one observation at a time. It is a less general technique than bootstrap, but it works better than bootstrap with complex samples.

Konrad also briefly discussed a number of non-bootstrap methods to estimate prediction error. These included apparent error, where the same data are used to fit the model and to test it; adjusted residual squared error, where several regressions are run using an unbiased estimator for variance; Mallows' Cp statistic, which works by plotting Cp against the number of variables +1 for alternative regressions using all possible combinations the variables;[25] and the Bayesian information criterion, which is based on residual squared error and a log n transformation. Unfortunately, all of these methods make assumptions that the number of parameters is known and the data are fit for the type of model used. In contrast, bootstrap and cross validation do not make such assumptions and work well even when the models being fit are incorrect. For example, using the plug-in principle, bootstrap can be used to estimate population parameters such as correlations in non-normal settings. Bootstrap methods have been used for analyses of the performance effects of CMM-based process improvement [16].

Prior to finishing his presentation, Konrad discussed a few applications of bootstrap and cross validation techniques that have been used in analyses of process performance data. He spoke most about classification and regression tree (CART) analyses, which use semi-automated methods that make extensive use iteratively of cross validation to construct a summary tree by splitting variable and values to best discriminate among the model outcomes. CART can process high-dimensionality data to create a decision tree that approximately optimizes prediction error. In fact, prior to the 1987 SEI technical report [15] that preceded the development of the original Software CMM, CART was used in unpublished work to help identify which questions would give insight to the original maturity levels. As few as five questions could distinguish among maturity levels. In addition, the team used cluster analysis to decide what would become key process areas. CART also has been used to analyze the performance effects of CMM-based process improvement [13]. Finally, Konrad touched briefly on capture-recapture methods that use resampling to estimate characteristics of small or difficult-to-reach populations. Originally developed in wildlife biology to monitor the census of bird, fish, and insect populations, capture-recapture methods have been used to optimize inspection decisions for largely defect-free but critical software [17].

3.2   Panel Session: Process Performance Models: Issues, and Obstacles

David Raffo led this panel discussion, basing his remarks on over two dozen process modeling projects done in multiple organizations at CMMI maturity levels 3, 4, and 5. Their experiences with process performance modeling have ranged from "wild success" and integration into decision making processes to just a project or two or not used altogether. Most of the organizations have completed multiple projects. Some have created small departments or other groups to do their process modeling. On the whole, success seemed to be positively correlated with how well the organization understood advanced methods and was willing to make decisions based on facts and data.

---

25   Cp is often used as a stopping rule for various forms of stepwise regression, which Mallows warned against doing.

Raffo began by listing several common challenges to the successful adoption and use of process performance models. For the kind of simulation-based work that he does, these challenges have included concerns about the initial cost in terms of both tools and people. As is so common elsewhere, he also mentioned difficulties in keeping management interest and support and in building trust in the models and their results. Several misconceptions he has encountered pertained to misunderstandings about the level of measurement and accuracy of data necessary to get useful results. Others had to do with how much commitment was required to be successful, with respect to high costs and likely benefits as well as the people involved. Similarly, perceived risks also have been common with respect to "untested technology," the credibility of the model results, credibility and availability of the necessary data, and the credibility of predictive models in general. Many of these challenges of course are not uncommon when implementing other aspects of process improvement as well as goal-driven measurement and analysis. And they can be exacerbated when communications are deficient among multiple stakeholder groups.

Several things have helped Raffo overcome such challenges. It is easier to articulate the likely benefits based on prior experience. Moreover, iterative development helps the modelers deliver beneficial results early and often. Securing management buy−in and commitment requires the modelers to stay focused on management's needs as well as ensure that the credibility of the data and results are in consonance with project goals. That is, find out what is giving people pain, and give them a way to address previously intractable or difficult problems. The ability of simulation-based models to quickly evaluate what-if tradeoffs can be a big win under such circumstances.

Raffo emphasized the importance of "one decision payback"—that is, do your best to ensure that you deliver valuable results early. Once the initial model is instantiated, the highest costs are amortized quickly and much more value can accrue at marginal cost. He also emphasized "consensus credibility." Good models can be a means for persuasion. If your audience doesn't agree with the assumptions or formulas in the model, substitute their preferences in the model and then compare the model results.

Kathy Smith spoke next, discussing the challenges that she has encountered with projects using TSP.[26] She began by noting that TSP quality profiles are used as leading indicators to identify the need to take corrective action, which maps well to the CMMI Quantitative Project Management process area. She also pointed out that the quality profiles can be used predicatively to evaluate and change processes. TSP re-launches provide an opportunity for changes that map well to the Causal Analysis and Resolution process area. In principle, they also can be used in the context of Organizational Innovation Deployment. As such, the quality profiles can be considered as elements of process performance models. It also could be straightforward for TSP practitioners to focus on modeling factors about how reviews are conducted.

Finding the necessary expertise to establish and fully develop process performance models is perhaps the major challenge that Smith has experienced in working on TSP projects. Expertise to build such models typically is not yet widespread. A related challenge has been providing motivation for useful analysis, not simply the production of charts. TSP practitioners could make better use of regression analysis and baselines to build predictive models. While TSP data does roll up to the team and organizational level, more attention may be needed to fidelity and consolidation of

---

[26] More details about the implications of PSP/TSP can be found in the synopsis of Smith's presentation materials in this report.

the individual data. It could be useful to collect consistently a small, efficient set of base measures to drive model-based decision making. Difficulties also have arisen due to off-shore operations and maintenance of legacy systems. That said, the potential clearly is there. TSP teams focus on team progress, quality, and productivity. They also typically focus on ensuring data quality, integrity, and granularity, and much of their data are collected and used weekly.

The third speaker was Pedro Colla. He too emphasized the importance of building models with the assistance of subject matter experts (SMEs). Doing so is necessary to ensure the model's credibility as well as validity among its intended users, some of whom are the SMEs themselves. The major obstacle or challenge is getting the SMEs and their managers to realize that it takes time to achieve added value, especially if they are short on time and resources. For similar reasons, it is more difficult to establish sufficient credibility to justify the adoption of complex models if there is a long lag time before results are available for use in decision making.

Mr. Colla also emphasized the importance of recognizing and communicating that while all models are wrong, some are useful. The model content and structure should be driven with a common-sense approach from the point of view of its intended users, management, and line engineers as well as the SMEs. The model is only as good as what it conveys to its users. On a related note, model builders must remain sensitive to the size of the organization and the time necessary before model results will contribute to the organization's process capabilities. Finally, Mr. Colla emphasized the need to take a structural view of how the models are built, verified, and validated. Doing so is both difficult and extremely important, especially in resource-poor environments.

Roz Singh was the last scheduled panel speaker. She discussed the fact that it can take a good deal of time to successfully introduce process performance modeling into a large and well-established organization. It took one organization she worked with 10 years' worth of data and 7 years to develop its first prototype PPM. They focused first on data quality and integrity, partly because of the existence of many long-duration projects. The emphasis on their baselines was of course important, but they paid much less attention to the models and how they should be used. The modeling team focused on statistical and mathematical methodological defensibility, but paid too little attention to the importance of domain knowledge from subject matter experts, stakeholder involvement, and the essentials of goal-driven measurement and analysis. For example, a few simulations may have facilitated faster progress, but the model team lacked the domain knowledge to do the necessary what-if scenarios.

In order to get statistical results in which they had confidence, the team spent years on data integrity and iteratively developing standardized operational definitions. For example, the organization did not have a common definition, as previously thought, of defect severity and what constitutes a major versus a minor defect. However, little attention was paid to the model's user interface. Moreover, the team did not recognize that successful adoption at the project level required simple models that were easy to use. Benefits at the organizational level are not enough. Models have to be useful for individual projects and programs too. Lacking stakeholder participation in developing the model, they had to repeatedly market and remarket it to the business and to each individual project. Eventually mandates were necessary to get the model into use.

The model eventually did provide useful information, but the going was tough. For example, after one program seemed to be doing too many peer reviews, the model was used to determine how many peer reviews were needed to minimize defects in later phases. However there had been a

great deal of variation due to project type, project size, duration, and domain, which remained unrecognized without the participation of subject matter experts. More segmentation than was anticipated was necessary to get a handle on variation in the measures used by different projects, which was problematic because of the resulting small data sets. In the end, the team was able to do some useful segmentation of the data, although they didn't have sufficient information to recommend what types of inspections ought to be performed.

Several other topics were raised both during and after the panel speakers concluded their remarks. For example, tradeoffs with respect to data quality between rigor and meaning first came up when Singh spoke about taking seven years to achieve confidence in using the data for statistical analysis. Several participants suggested that it would have been useful to consider discarding much of the old data. Some data are so bad they need to be discarded. Some decisions require very accurate data. However, a number of the workshop participants emphasized that useful analysis can be done with imperfect data as long as the variability is recognized in interpreting the results, for example through sensitivity analyses. Discussion ensued about how much granularity of the data and subprocesses is enough to justify analysis to gain sufficient insight for management or engineers to take action.

The need for domain expertise was raised repeatedly. However, domain experts need to understand what the modelers need to know too. Brook Eiche suggested that a three-person core PPM team often can be a big help, including a subject matter expert, modeling expert, and a Six Sigma black belt to tease out the purpose and goals. The key appears to be building something small with value added to get others to ask for more of the same. Start with small scope and build on it after asking the team members and others to discuss what keeps them up at night and identify pertinent factors and outcomes that might benefit by modeling. Using this approach, Eiche's group was working on seven models with another twenty on the horizon at the time of the workshop.

Other pertinent topics that were discussed are listed below.

- Models that outperform the experts can be particularly convincing.
- Environments with a large proportion of older workers can face cultural challenges when introducing and using models; however, changes in the workforce also can affect which factors will be significant.
- Business objectives need to drive PPMs, but a single objective may drive multiple PPMs.
- Model developers need to be sensitive to the costs of data collection to determine whether to include various x factors in a model. Statistical relationships and relationships to important business goals also are important criteria to be used in deciding what to collect.
- Statistical significance does not equate with practical significance.
- Issues remain about tradeoffs between collecting coarse-grained data versus regretting when you do not have sufficient granularity when you need it.
- Collecting a lot of data that no one uses is not a good way to achieve credibility for the modeling effort.
- The need for exploration and understanding as distinguished from routine reporting should be recognized.
- A census mentality should be avoided. Sampling can be better as well as cheaper.
- The analysis must add value. The analytical methods are a means, not an end.

Finally, the topic of what is, or should be, meant by a CMMI-based process performance model was brought up by other workshop participants during the panel.

It was first pointed out that the definition remained poorly understood. Others agreed, and some dissatisfaction with the appraisal process was expressed. Still others responded that it was crucial to have contextual information about the model and how it is used to judge whether it does or does not qualify as a CMMI PPM. Usage is a more important factor than strict rules in determining what should be considered a PPM. CMMI Version 1.3 needs to provide more clarification about using quantitative thinking to recognize a mix of acceptable approaches to PPMs and PPBs. More guidance is needed to identify what should be encouraged as well as what should be discouraged. The role of CMMI is to provide useful guidance that is valuable for its users. Compliance is not an end in itself. This may be a model issue, not simply one of appraisal.

Regardless, modelers need to be better briefed on what is expected for CMMI-based PPMs. To that end, black belt training at Lockheed Martin and elsewhere now includes extended training and orientation on process performance modeling. Some discussion also ensued about the healthy ingredients of process performance models. More clarification may be valuable about interpreting them for different types of PPMs and the purposes for which they are used. It also might handle some of the misconceptions about having "enough data" for a PPM.

3.3 Break-Out Sessions

Use and Outcomes of Process Performance Modeling

Participants in this workshop session focused on how best to summarize the use and outcomes of process performance modeling and make the results available more widely throughout the CMMI-based process improvement community.

When topics were being discussed for the breakout sessions during the workshop, particular interest was expressed in the importance of sharing information about collecting and summarizing case study information about ROI and NPV. However, the scope of the discussion was wider during the session. The group recognized that claims of value added by process performance modeling had to be augmented with a fuller description of the modeling methods as well as how and why the models were created and used. The focus and level of detail would need to differ, but more information would be needed whether the intended audience was familiar with PPMs and PPBs or just getting started.

The breakout group discussed several topics. They began by discussing what had to be made clear about the meaning of a process performance model in CMMI. What steps are necessary to create and deploy a PPM? What is, or should be, the design and architecture for modeling an entire system? What is an agreed-upon CMMI definition of a PPM? The group agreed that the presentation template used for the workshop presentations was a good place to start in explaining these issues, as was an explicit link to CMMI model practices.

The group also discussed the importance of describing the kinds of analytical methods and techniques that can be used for process performance modeling. A catalogue of such methods might be a useful approach. What are the pros and cons of the different methods for use in a PPM? What

sorts of resources are needed to support such work? What are the data requirements and assumptions?

The group next addressed the kinds of training and mentoring that would be helpful in institutionalizing the use of PPMs and PPBs. Although role boundaries vary, what is needed clearly differs depending on the backgrounds of the stakeholders. Managers and line engineers need to understand the results and learn to frame their questions about what kinds of results the modelers should provide. The same is true for other model users and analysts. Those who build models clearly need more detail, but what and how much they need to know depends on their initial levels of expertise. Software developers and model developers need to create appropriate architectures to build tools to support the models. What kinds of mentoring and other support systems exist for the modeling? How is model use reinforced and diffused throughout the organization? What about CMMI model expertise (e.g., the generic practices being considered)?

Management sponsorship was another topic that the group discussed. What does it take to establish credibility for doing process performance? Focusing solely on getting a high maturity level is counterproductive. What is a good business case for investing in model-based prediction capability? What is the cost of developing that capability? What is the marginal investment beyond current state required to achieve an initial modeling capability? More case studies are needed. Can the workshop participants create them? Do other pertinent case studies already exist (e.g., in the Six Sigma venues)? Can we replicate the studies and their results?

The group concluded by discussing the kinds of evidence that needs to be collected and summarized to determine the value of doing process performance modeling. How is the performance of a model evaluated in business terms, whether expressed in ROI, NPV, product quality, fitness for use, or any other criteria? There is a need to go beyond statistical measures of fit. Do the model outcomes and the guidance based on them lead to better business performance?

Approaches to Process Performance Modeling

A wide range of topics were discussed in this session. The participants began by discussing several questions about the scope of topics that can be usefully addressed by process performance models. These included aspects of organizational processes. For example hardware development and test could be modeled to evaluate the impact of change across projects with respect to quality and testability. Program staff effectiveness could be evaluated using controllable and uncontrollable x factors such as defined roles and responsibilities, staff turnover, interruptions, and productivity. Do we have the right team assembled to do the particular job? Is attrition affecting capability to perform defined processes? Similar questions were raised at the project level. How can we better meet schedule and understand the factors that affect it? The group also discussed interim outcomes. How can we better evaluate the impact of in-process changes in specific projects? Can we predict whether or not such changes would help a project meet schedule? Would changes in testability or test coverage affect predictions of failure rate in design?

The group also discussed several deliverables that they thought would help progress the state of the practice of process performance modeling. These included the preparation of detailed examples of a variety of models including possible outcomes and controllable and uncontrollable x factors, along with templates describing each in detail. Such templates could include such things as business usage, modeling techniques, model utilization, data collection, and example operational

definitions of outcomes and x factors. The discussion also focused on the importance of including more guidance about handling contextual information such as the association among x and y factors with explicit subprocesses, other aspects of the nature of the business being modeled, the technology being used, size of the organization, intricacy of the product being developed, investment required, and opportunity costs.

The Future: "Goal-Question-Model"

A small group discussed what was necessary to provide useful guidance to modelers and the wider community about using process performance models to improve business and technical performance.

Confusion, misunderstanding, and disagreement still exist about what qualifies or should qualify as a CMMI-based process performance model and how best to provide actionable guidance to those who do or should do process performance modeling. More guidance clearly is needed for modelers as well as the wider community, yet the discourse too often focuses on notional ideas without sufficient context. The group considered areas that need to be addressed for the community to converge on a shared vision of what adds value.

Akin to GQM as applied to measurement and analysis in general, one of the group members suggested that a similar approach be taken in the session to stay focused on what it is that they were trying to accomplish. The "M" in GQM could be considered shorthand for process performance model.

Whether for management, line engineers, or the modelers themselves, the guidance must be both actionable and credible if it is to be understood and followed. Notional assertions of best practice must be backed up by practical examples to be meaningful to people who are not already familiar with the topic. As one of the group participants said, it is "wacko" to do what may not be valuable for you.

Real-world examples are absolutely necessary to demonstrate convincingly that using models can and has improved business and technical performance. Such examples must be made available from a wide variety of industry and government settings. That is true whether the focus of the examples is on the genesis and makeup of the models, the model outcomes, or the results of having used them. Such examples can be disseminated via publications, presentations, training, or mentoring, whether via paper, electronically, or face-to-face. Within reason, the results can be sanitized or normalized, but explicit values must be seen if value is to be added by the guidance.

Commonalities of successful model building must be shared more widely throughout the community in much the same way, focusing on the purposes, processes, and resulting model structure in sufficient detail to be understood by the intended audience. Model builders clearly will benefit from learning about analytical techniques that have been proven useful. The same is so—albeit with less detail—for management and other important stakeholders. All of the stakeholders need to know more about helpful adoption and deployment strategies, perhaps especially the modelers.

In addition, the group discussed what needs to be done beyond and between these workshops. Sharing experience and insights in the workshops has been invaluable for many participants. The group agreed that it is important to maintain constancy of the core members over time, but that

must be balanced by adding more people and organizations as well. The workshop participants benefit by the self-help experience, but they must share what they learn with others. The group also agreed that the results of these workshops and the work they bring about need to be provided proactively to influence the evolution of CMMI models, appraisals, and training.

Deployment and Adoption of Process Performance Models

Another group discussed a wide range of issues having to do with both the deployment and adoption of process performance models. Much of the conversation centered around anticipated and desired changes to Version 1.3 of the CMMI models. There was some related discussion about the limitations of earned value management as well as the synergy of TSP and CMMI.

The group began by agreeing that there is confusion in the community about what constitutes a valid CMMI process performance model. As the PPM name implies, everyone realizes that high maturity needs to be about performance. But what differentiates maturity levels 4 and 5 from level 3, and what is the role of PPMs? While there clearly are similarities in coverage and goals, the analytical approaches differ considerably. For example, the group contrasted Decision Analysis and Resolution (DAR) and Integrated Project Management (IPM) at maturity level 3 with Causal Analysis and Resolution (CAR) and Organizational Innovation and Deployment (OID) at maturity level 5. The criteria for meeting the goals of maturity level 3 can be fixed, static, and deterministic. At the higher maturity levels, the emphasis becomes more focused on understanding relationships with their root causes and recognizing the importance of objective analysis under conditions of uncertainty. The purpose of modeling goes beyond estimation to explanation and understanding relationships, including those that can improve estimation accuracy. The group also agreed that the higher maturity levels differ by giving more direction about how to improve as well as what to improve.

The group agreed that Version 1.3 should put more expectations of what constitutes high maturity practices into the models' expected material rather than the informative material. Recognizing that it would be counter-productive to change the models too much for Version 1.3, the group also suggested that the models be more explicit about what enables successful adoption of high maturity practices as well as what constitutes such practices. The move to Version 1.3 provides the community with an opportunity to identify gaps in the community's understanding of high maturity practices in general and process performance modeling in particular.

During the breakout session, the group also started building a table to outline what is necessary for various stakeholder groups to understand the meaning of high maturity. The group suggested that the Version 1.3 models and ancillary publications be more explicit about addressing the points of pain and other needs in situations that were likely to arise for different stakeholders. For example, at a minimum, pertinent examples for engineers would discuss issues of cost, schedule, and quality that relate to products and processes. Similar issues rolled up across organizational units would be more compelling for management in general. Senior management could benefit from examples about behaviors necessary to enforce attention to cost, schedule, and quality. Variants on these themes could be useful for others such as bid managers.

Guidance about process performance models should also vary by situation. For example, engineers could benefit by more specific guidance about building and validating models incrementally. Examples about how PPMs can be used as critical analytical tools for their own work and ex-

amples from TSP core models might help. However, models that are more useful for management can differ considerably. For example, they might focus on managing differences across time zones and stakeholder organizations, both horizontally and vertically across organizational units. The focus might be more strategic than tactical and reflect greater conditions of uncertainty. Similarly, guidance about measurement and analysis can vary across stakeholders. For project or program engineers, data elements may be quite small and granular at the level of subprocesses. They also are likely to be important base measures for models that are more pertinent for other stakeholders.

The group emphasized that more explicit examples are needed about the successful adoption of process performance modeling. How are such models "born"? Is there a passionate owner of the model? Was an important stakeholder experiencing pain that the model was created to diminish? Regardless of the stakeholder group, the focus for those who have not already used PPMs successfully should be on to what end and how a PPM can be used valuably. An initial focus on the analytics and data structure can be entirely counter-productive. The point to be emphasized should be how valuable the PPM can be for its intended users. Some organizations will already have models that fully satisfy CMMI criteria for a PPM. Others may have constructs that can easily become full-fledged PPMs. Yet, while explaining how and why such models can be extremely useful, it well may be the better part of valor in other organizations to call them something other than PPMs.

Other important issues that the group emphasized should be made more explicit with respect to the successful institutionalization of PPMs as well as their adoption are listed below.

- Process performance models should be tied closely to work being done by the teams, regardless of the stakeholder groups involved.
- The models need to be tools for analysis that help with critical work activity.
- As such, the measures used in the modeling should be rooted in granular work activities.
- The measures can be, and many should be, sufficiently finely grained to support process improvement that results in better performance and quality throughout the product life cycle, such that adjustments can be made early if and as necessary.

Analytical Methods, Techniques, and Tools

This breakout group discussed the adoption issues of process performance modeling with respect to the barriers inherent in the analytical methods, techniques, and tools. The group decided to brainstorm the different barriers and issues so that a subsequent prioritization would enable an off-line working group to make progress on solutions to help the community. Barriers and issues are explained below and represent prime topics for future workshops.

1. Interest in application to CMMI for Services vs. Acquisition vs. Development
   A need exists to specify and share examples of process performance modeling methods, techniques, and tools appropriate within a CMMI for Services high maturity setting, to include possibly unique applications of process performance modeling that would be quite different from the acquisition and development settings.
2. Application space
   A need also exists to demonstrate and share examples across various software and systems applications (e.g., defense, IT, medical, financial, and manufacturing).

3. Method space
   It would help if the working group could document and share the landscape of practical analytical methods available to the community as a first step in addressing the issue of organizations "not knowing what they do not know."

4. Job aids
   Job aids are needed to support decisions on which analytical technique to use in a given situation and for how to combine the use of different techniques.

5. Guidance, not dogma
   A documented process to build PPMs would be invaluable to community adoption.

6. DAR criteria for PPMs
   An idea surfaced regarding the value of creating a template DAR criterion in the high maturity space to help organizations think through the specific decisions related to process performance model usage.

7. PPM usage profiles
   A collection of potential use cases, including a description of the use case activities and their sequence related to process performance models would help organizations plan the role of PPMs in the organization and project operations.

8. Tool comparisons
   Some variant of a *Consumer Reports* type of presentation of information with regard to the different commercial tools would be helpful for organizations deciding which analysis and modeling tools to invest in.

9. Feedback and testimonials
   Adopters of CMMI high maturity and process performance models would greatly benefit from the establishment of a mechanism for feedback/testimonials for tools and experience.

10. Community sharing
    One way to encourage community sharing of experience would be something similar to a CMMI high maturity Wikipedia, along with the creation of a template of information to use. The governance model for the Wikipedia operation would need to be considered.

11. Case study template
    It would help to document and publicize a case study template for the development, use, and overall experience of process performance models, along with a sharing mechanism.

12. Common business problems and goals
    Some breakout group members felt it would help to create an affinity of items associated with the nature of the common business problems and related business objectives. A possible short list of such problems and example business objectives would further ease the initial activity by organizations new to planning CMMI high maturity activities such as process performance modeling.

13. PPM views or filters
    An additional suggestion during the breakout group discussion was that many views or filters would help. They could be based on organizational function, discipline (including hardware vs. software), program, nature of outcome predicted, need, model use, logical, physical, functional management, product management view, improvement management view, and so forth.

14. Prescriptive guidance
    Applied guidance (experiential) related to the use of analytical methods for PPMs should be

considered and made available to the community, possibly as a wizard-type tool developed to add both managers and model builders.

15. PPM healthy ingredients

    There is a need to refine and clarify the healthy ingredients for process performance models and baselines.

16. References

    An authoritative list of texts, references, articles, sources of training, and training itself needs to be shared with the community. Specifically, online training development should be prioritized for development by the SEI and others.

17. Data sets

    The breakout group cited the opportunity and benefit of collecting and sharing practice data sets and solutions for different tools and modeling techniques. This would support training and coaching, especially for small- to medium-sized organizations.

18. Current awareness

    It was suggested that the group establish an awareness program on modeling techniques, tools, and achievements by enlisting owners of different topics.

19. Publications

    A continuous stream of publications on process performance modeling and results could be organized, with publications used by the CMMI community.

20. Monthly teleconference calls

    Progress on this list of potential actions would be accelerated through the use of monthly teleconference calls occurring between the semi-annual CMMI High Maturity Measurement and Analysis workshops.

Since the third workshop, some of these actions, listed below, were accomplished.

- The SEMA team has made progress in evaluating the CMMI Acquisition and Service spaces for unique opportunities to use process performance modeling,

- Several SEMA presentations and tutorials at SEPG conferences have sought to communicate the landscape of methods for process performance modeling. The SEI Understanding CMMI High Maturity Practices course[27] also communicates this practical landscape of methods.

- The SEMA team has created numerous job aids for decision making in the use of different analytical methods toward process performance modeling. What remains absent are the criteria on which methods should be explored first given a specific situation.

- This technical report sets a precedent in terms of SEI publications seeking to share a compendium of real world case studies of the actual implementation and results of process performance modeling.

---

[27] Course information is available on the SEI website at http://cmsauth.sei.cmu.edu/training/p14b.cfm.

4 Summary and Next Steps

By making the models that were shared in these two workshops more widely available in this report, we hope that the community as a whole will benefit from the exciting and innovative ideas for process performance models implemented by leading organizations in the field. As shown in this report, these organizations have developed and used many different analytical approaches in their modeling efforts. Most of them have used statistical and probabilistic methods. Several of the models have been what-if simulations. A few also have incorporated deterministic mathematical modeling techniques. Many have used large-scale baselines, while others have relied on smaller datasets. Some also have addressed issues of data quality and integrity, and expert judgment has been used in the absence of other data to calibrate Monte Carlo models and discrete event simulations.

Most of the presentations described modeling experiences in large organizations consisting of multiple and sometimes disparate stakeholders. However, some of the presentations showed how disciplined modeling approaches can and have been used in smaller organizations. Of course the interim and final performance outcomes predicted by the models differ considerably as a function of the business and technical objectives that the models were built to help achieve. And the models' predictive factors differed as a function of the uncontrollable business and technical constraints faced in meeting those objectives as well as the process and negotiable factors that were under the control of the models' users.

As seen in the presentation synopses, many kinds of performance and quality outcomes were predicted by the models discussed during the workshops. The presentations focused on many things, including various aspects of removing and preventing defects; measures of customer satisfaction and other quality attributes; and the predictability and magnitude of resource costs and schedule. Others focused on things like cost of quality, ROI, or staff capabilities. They all described the use of process performance models to predict, manage, and improve controllable factors and their resultant business outcomes. Many of the results were quite notable and used effectively in decision making.

As noted in the introduction to this report, the broader goal of this workshop series is the establishment of a viable community of interest around high maturity measurement and analysis. We already are seeing considerable improvement in the technical quality and increased application of process performance models in the CMMI-based process improvement community. It was evident most recently in the number of proposals focusing on high maturity measurement and analysis practices that were submitted and accepted for presentation at the 9th annual CMMI Technology Conference and User Group in November 2009. Similarly, we have been pleased and even a little surprised at the results we have seen in our recent surveys on the state of high maturity measurement and analysis practices [18].

A fourth workshop in this twice-yearly series took place immediately following the 2009 CMMI Technology Conference. Over 20 deserving presentations were accepted for presentation there,

which is more than could fit workshop spanning two full days. As a result, several presentations are taking place between workshops via teleconference and the internet.

Other activities between workshops are being considered at this writing to encourage more sharing and collaboration among what is fast becoming a larger community that can no longer rely solely on face-to-face contacts. More must be done to better accommodate this growing community of peers and make the work more widely known. We need mechanisms to publish results incrementally, most likely via the web and journals as well as traditional reports such as this one. In order to make that possible, a large group of peers is needed to review each other's work to their mutual benefit. There also clearly is a need to limit full access to the workshop presentations and their results to active participation among those at the cutting edge, but that too can be done remotely.

We also will be broadening the scope of future workshops and publications to include more descriptions and lessons learned about issues of adoption, implementation, and management buy-in. Descriptions of uncommon approaches to data quality and integrity such as voice of the customer and other "soft side" Six Sigma approaches also may be included along with other useful measurement, modeling, and analytical approaches that may not yet be incorporated in process performance baselines and models. A guide for developing process performance models is also being developed. All of this we hope and trust will lead to better use of measurement and analysis in support of processes and technologies that enable the development, sustainment, and acquisition of products and services that meet their intended use and are available on a timely and cost-effective basis.

References/Bibliography

URLs are valid as of the publication date of this document.

[1]   R. Stoddard, D. Goldenson, D. Zubrow, and E. Harper, "*CMMI High Maturity Measure-
        ment and Analysis Workshop Report: March 2008,"* Software Engineering Institute, Car-
        negie Mellon University, 2008, Tech. Rep. CMU/SEI-2008-TN-027.
        http://www.sei.cmu.edu/library/abstracts/reports/08tn027.cfm

[2]   J. Liker and D. Meier, *The Toyota Way Fieldbook*. New York, NY, USA: McGraw-Hill,
        2006.

[3]   W. Harrison, J. Settle, and D. Raffo, "Assessing the Value of Improved Predictability due
        to Process Improvements," Third International Workshop on Economics-Driven Software
        Engineering Research (EDSER-3 2001), Toronto, Ontario, Canada, May 2001.

[4]   P.K. Lawlis, R.M. Flowe, and J.B. Thordahl, "A Correlational Study of the CMM and
        Software Development Performance," *Crosstalk: The Journal of Defense Software Engi-
        neering*, September 1995, pp.21-25.

[5]   P. Colla, "Gestión de Mejoras en Procesos en Ingeniería de Software con el Soporte de
        Modelos Sistémicos: Aplicación al Sector de Desarrollo de Software de Argentina," Tesis
        de PhD, Universidad Tecnológica Nacional-Facultad Regional Santa Fé (UTN-FRSF),
        2009, ISBN: 978-987-05-8108-6.

[6]   P. Carbonell-Foulquié, J. Munuera-Alemán, and A. Rodríguez-Escudero, "Criteria Em-
        ployed for Go/No-Go Decisions When Developing Successful Highly Innovative Prod-
        ucts," *Industrial Marketing Management*, vol. 33, pp. 307-316, August 2003.

[7]   A. Rosene, J. Connolly, and K.M. Bracu. "Software Maintainability: What It Means and
        How to Achieve It," *IEEE Transactions on Reliability*, vol. R-30, no. 3, August 1981.

[8]   Y. Luo, T. Bergander, and A. Ben Hamza, "Software Reliability Growth Modeling Using
        a Weighted Laplace Test Statistic," *Compsac*, vol. 2, pp.305-312, 2007, 31st Annual Inter-
        national Computer Software and Applications Conference, 2007.

[9]   W. Farr, "Chapter 3, Software Reliability Modeling Survey of *Handbook of Software
        Reliability Engineering*," M. Lyu, Ed.: EEE Computer Society Press, 1997.

[10]   P. Luo Li, M. Shaw, J. Herbsleb, B. Ray, and P. Santhanam, "Empirical Evaluation of De-
        fect Projection Models for Widely Deployed Production Software Systems. "ACM
        SIGSOFT Software Engineering Notes, vol. 29, pp. 263-272, November 2004.

[11]   S. Kan, Chapter 7 in *Metrics and Models in Software Quality Engineering, Second Edition*,
        Boston, MA: Addison-Wesley Professional, 2002.

[12]   B. Efron and R. Tibshirani, "Chapter 17: Cross-Validation and Other Estimates of Predic-
        tion Error" in *An Introduction to the Bootstrap*, Chapman & Hall LTD., 1993.

[13]   Project Management Institute, *A Guide to the Project Management Body of Knowledge, Fourth Edition.* Newtown Square, PA: Project Management Institute, 2008.

[14]   H-W. Jung and D. Goldenson, "Evaluating the Relationship Between Process Improvement and Schedule Deviation in Software Maintenance," *Information and Software Technology*, vol. 51, issue 2, February 2009.

[15]   W. Humphrey and W. Sweet*, "*A Method for Assessing the Software Engineering Capability of Contractors," Software Engineering Institute, Carnegie Mellon University, 1987, Tech. Rep. CMU/SEI-87-TR-023.
http://www.sei.cmu.edu/library/abstracts/reports/87tr023.cfm

[16]   K. El-Emam, D. Goldenson, J. McCurley, and J. Herbsleb, "Success or Failure? Modeling the Likelihood of Software Process Improvement," *Empirical Software Engineering*, vol.6, issue 2, 2001.

[17]   J. Barnard, K. El Emam, and D. Zubrow, "Using Capture-Recapture Models for the Reinspection Decision," *Software Quality Professional*, vol. 5, issue 2, pp. 11-20, 2003.

[18]   D. Goldenson, J. McCurley, and R. Stoddard, "Use and Organizational Effects of Measurement and Analysis in High Maturity Organizations: Results from the 2008 SEI State of Measurement and Analysis Practice Surveys," Software Engineering Institute, Carnegie Mellon University, 2009, Tech. Rep. CMU/SEI-2008-TR-024.
http://www.sei.cmu.edu/library/abstracts/reports/08tr024.cfm

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, search-ing existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regard-ing this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE January 2010	3. REPORT TYPE AND DATES COVERED Final
4. TITLE AND SUBTITLE Approaches to Process Performance Modeling: A Summary from the SEI Series of Workshops on CMMI High Maturity Measurement and Analysis		5. FUNDING NUMBERS FA8721-05-C-0003
6. AUTHOR(S) Robert W. Stoddard II & Dennis R. Goldenson		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213		8. PERFORMING ORGANIZATION REPORT NUMBER CMU/SEI-2009-TR-021
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116		10. SPONSORING/MONITORING AGENCY REPORT NUMBER ESC-TR-2009-021
11. SUPPLEMENTARY NOTES		
12A DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited, DTIC, NTIS		12B DISTRIBUTION CODE

13. ABSTRACT (MAXIMUM 200 WORDS)

More and more organizations are striving for and achieving high maturity status, yet there is still an insufficient shared understanding of how best to implement measurement and analysis practices appropriate for high maturity organizations. A series of twice-yearly work-shops organized by the Software Engineering Institute (SEI) allows organizations to share lessons learned to accelerate the adoption of best measurement and analysis practices in high maturity organizations.

This report summarizes the results from the second and third high maturity measurement and analysis workshops. The participants' presentations described their experiences with process performance models; the goals and outcomes of the modeling; the x factors used; the data collection methods; and the statistical, simulation, or probabilistic modeling techniques used. Overall sum-maries of the experience and future plans for modeling also were provided by participants.

This report also includes a summary of the "healthy ingredients" that are needed for process performance models and a table showing which healthy ingredients were visible in the models de-scribed in the presentations. By making the models that were shared in these workshops more widely available in this report, the community as a whole can benefit from the exciting and innovative ideas for process performance models implemented by leading organizations in the field.

14. SUBJECT TERMS Process performance model, healthy ingredients, CMMI, high maturity, SEMA		15. NUMBER OF PAGES 119
16. PRICE CODE		

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18
298-102